مجله مهندسی مالی و مدیریت اوراق بهادار
شماره نوزدهم / تابستان ۱۳۹۳

بهینه سازی سبدسهام با استفاده از روش k-means و الگوریتم ژنتیک

ابراهیم پورزرندی[۱]
مینا کیخا[۲]

تاریخ دریافت: ۹۲/۱۰/۱۷ تاریخ پذیرش: ۹۳/۱/۲۰

چکیده

دیدگاهی که در این مقاله ارائه می دهیم در دو مرحله جای می گیرد: مرحله ی اول طبقه بندی سهم ها ی پورتفوی ابتدایی با روش k-means به دسته های کوچکتر است، سپس طبقه ای که کمترین ریسک و بیشترین بازده را دارد یا به عبارتی طبقه ای که بهینه تر می باشد را به عنوان ورودی الگوریتم خود که آن را MinVaRMaxR نامیده ایم برمی گزینیم. الگوریتم مذبور،الگوریتم پویایی، براساس الگوریتم ژنتیک و مفهوم ارزش در معرض خطر می باشد. هدف ما از اجرای این الگوریتم، مینیمم کردن ریسک و ماکسیمم کردن بازده پورتفوی، به صورت همزمان می باشد. به همین منظور ۱۰۰ شرکت عضو بورس اوراق بهادار تهران طی سالهای ۱۳۸۸، ۱۳۸۹ و ۱۳۹۰ به صورت روزانه مورد مطالعه قرار گرفته است. توزیع داده ها نرمال نبوده، به همین دلیل از آزمون های آمار ناپارامتری جهت آزمون فرضیات استفاده شده است.

نتایج تحقیق بیانگر آن است که طبقه بندی داده ها وسپس اجرای الگوریتم ژنتیک روی طبقه ی بهینه موجب دستیابی به پورتفویی می شود که نسبت به پورتفوهای حاصل از اجرای ژنتیک به تنهایی، دارای ریسک کمتر و بازدهی بیشتر است. همچنین در مورد سبدهای با اندازه ی کوچکتر، الگوریتم ژنتیک خود به خود سبدی را بر می گزیند که در طبقه ی بهینه ی انتخابی الگوریتم طبقه بندی جای گرفته است.

واژه‌های کلیدی: الگوریتم ژنتیک، اگوریتم طبقه بندی، روش k-means ، ارزش در معرض خطر.

۱- دانشیار گروه مدیریت صنعتی دانشگاه آزاد اسلامی واحد تهران مرکزی pourzarandi@yahoo.com
۲- کارشناس ارشد مدیریت صنعتی دانشگاه آزاد اسلامی واحد تهران مرکزی minakeikha67@gmail.com

۱- مقدمه

در کشور های توسعه یافته اکثر سرمایه گذاری ها از طریق بازارهای مالی انجام می پذیرد. مشارکت فعال افراد در بورس متضمن حیات بازار سرمایه و توسعه پایدار کشور است.[۲] چرا که اگر پس اندازهای افراد با مکانیسم صحیح به بخش تولید هدایت شود، علاوه بر بازدهی که برای صاحبان سرمایه به ارمغان می آورند، می توانند به عنوان مهم ترین عامل تأمین سرمایه برای راه اندازی طرح های اقتصادی جامعه، مفید باشند.[۱۰] سوال این پژوهش، سوالی است که هنگام ورود به بازار سرمایه، سرمایه گذاران با آن مواجه هستند وآن مسأله ی تصمیم گیری برای انتخاب و تشکیل سبد بهینه ای از سهام می باشد، چرا که خود را در مقابل گزینه های بسیاری می بینند و باید انتخاب کنند که کدام سهم شایستگی قرارگرفتن در سبد سرمایه گذاری را دارد، که برای حل این مسأله نیازمند بکارگیری ابزارهایی جدید جهت کسب سود بیشتر هستند.

هدف این پژوهش نیز بر همین مبنا انتخاب سبد سهامی از میان سبدهای مختلف سرمایه گذاری است، به گونه ای که بیشترین بازده و کمترین ریسک را به دنبال داشته باشد و این هدف با به کار گیری الگوریم های طبقه بندی و MinVaRMaxR که این الگوریتم بر مبنای الگوریتم ژنتیک و ارزش در معرض خطر یا همان VaR است، محقق می شود.

مسأله ی انتخاب سبد بهینه برای نخستین بار توسط مارکویتز مطرح شد. وی ثابت کرد تنوع بخشی در انتخاب سهام و تشکیل سبد، ریسک آن را برای سرمایه گذار کاهش می دهد.[۷] اما دلایل به کارگیری هرچه بیشتر الگوریتم های ابتکاری به جای مدل مارکویتز در مسائل کنونی انتخاب پورتفوی، در این است که در حل مدل مارکویتز از راه حل ها و مدل های ریاضی استفاده می شود؛ مشخص است که مدل های ریاضی بسیاری از محدودیت های جهان واقعی را در بر نمی گیرند، مثل تعداد زیاد سرمایه یا مقادیر وزنی سهام. این عوامل که اضافه می شوند فضای جست و جو، هم بزرگ می شود و هم ناپیوسته، در نتیجه در عمل، استفاده از مدل ریاضی امکان پذیر نیست. اگراین محدودیت ها را نادیده بگیریم، قطعاً جواب های ما به دنیای واقعی نزدیک نخواهد بود؛ الگوریتم های ابتکاری در این جا جایگاه می یابند، چراکه این الگوریتم ها از طبیعت گرفته شده و قادرند این محدودیت ها را در بر بگیرد.[۹]

همانطور که بیان شد معیار سنجش ریسک در این پژوهش ارزش در معرض خطر یا VaR می باشد؛ در روش VaR اصول کار، برای انتخاب سبد سرمایه بهینه، شبیه مدل مارکویتز است، با این تفاوت که در مدل VaR هر ریسکی کاهش نمی یابد، بلکه ریسکی کاهش می یابد که منجر به افزایش بازده نگردد. برای همین است که بازده این روش بیشتر است.

$MinZ = Z\alpha\ \delta_p - \bar{r}_p$ (VaRمدل) $MinZ = \delta_p^2$ (مدل مارکویتز)

St: $\bar{r}_p = \sum_{j=1}^{n} w_j \cdot \bar{r}_j$
$\sum_{j=1}^{n} w_j = 1$
$W_j \geq 0$

St: $\bar{r}_p = \sum_{j=1}^{n} w_j \cdot \bar{r}_j$
$\sum_{j=1}^{n} w_j = 1$
$W_j \geq 0$

δ_p^2: واریانس بازده سبد سرمایه، $\bar{r}_p$:بازدهی مورد انتظار سبد سرمایه،W_j : وزن مربوط به سهام j در سبد سرمایه، $\bar{r}_j$:بازدهی سهم jام

همچنین توزیع نرخ بازده دارایی ها در مدل مارکویتز حتماً باید نرمال باشد و اگر دارای چولگی باشد، مدل مارکویتز کاربردی نیست اما VaR برای ابزارهای مالی که دارای توزیع نرمال یا غیر نرمال باشند، نیز کاربرد دارد. در نتیجه از مدل VaR به جای مدل مارکویتز استفاده شد.

۲- مبانی نظری و مروری بر پیشینه پژوهش

مساله انتخاب سبد بهینه برای نخستین بار توسط مارکویتز مطرح شد با این هدف که به سوالات زیر پاسخ دهد:

اگر کسی بخواهد مبلغی را اختصاص دهد تا در بازارهای مالی سرمایه گذار محسوب شود ، ترکیب (ساختار) این سرمایه گذاری باید چه باشد؟ چه ترکیباتی باید انتخاب شود و به چه نسبتی؟[۱۰] انتقادها ی زیادی به این نظریه با داشتن تابع هدف درجه دوم و محاسبه ماتریس واریانس-کواریانس که بسیار مهم است، شده است.جهت ساده تر کردن مشکلات مدل مارکویتز ، مدلهای بسیاری به عنوان مدلهای جایگزین ارائه شده است.

شارپ و استون[۱] تلاش کردند تا مساله ی انتخاب پورتفوی را به صورت خطی درآورند. رود و روزنبگ[۲] نشان دادند که مدل مارکویتز در فرمولبندی کلاسیکش هنوز از جایی که سرمایه گذار حرفه ای را راضی کند دور است و آنها مدیریت پورتفوی را ارائه دادند.

مدلهای برنامه ریزی خطی زیادی برای انتخاب پورتفوی توسعه داده شده است . کونو و یامازاکی[۳] ، زینوس و کنگ[۴] و همچنین اسپرانزا[۵] جهت محاسبه ی ریسک پورتفوی ، مفهوم انحراف معیار را ارئه دادند. همچنین آنها مدلی را ارائه دادند که معیار ریسک نامتقارن (بی قرینه) را در بر می گرفت که مشکلات زیادی را در مورد مدل بهینه سازی بر طرف می کرد.[۱۶]

در توضیح ریسک نا متقارن لازم است به مطلبی اشاره شود ؛ استفاده از واریانس برا ی یک دارایی که دارای توزیع نرمال باشد و در بازاری کارا معامله شود معیار قابل قبولی است. حال اگر این دو خصوصیت برای دارایی وجود نداشته باشد استفاده از واریانس با مشکل رو به رو می شود. به این دلیل معیار دیگری برای ریسک مطرح می شود که نیمه واریانس از جمله ی آنهاست. با توجه به این معیار تنها بازده های تصادفی که از میانگین بازده پایین تر باشند، در محاسبه ی ریسک مورد استفاده قرار می گیرند . در حقیقت در این تعریف از ریسک، میزان انحراف از بازده مورد

انتظار تا جایی خطر آفرین است که به زیان سرمایه گذار بیانجامد و در غیر این صورت انحراف از بازده هیچ گونه ریسکی ایجاد نمی نماید. لذا در محاسبه ی ریسک در مواقعی که مقدار بازده تصادفی، از بازده مورد انتظار بیشتر باشد ، صفر را جایگزین تفاوت آن دو می نماییم. [۲]

یوشیموتو[۶] سیستم بهینه سازی را بر مبنای معیار نیمه واریانس مطرح کرد. همچنین همزاو جانسن[۷] مدل برنامه ریزی خطی جامعی بر مبنای روش های برنامه ریزی جدا پذیر، ارائه دادند. اخیراً مقیاس جدیدی از ریسک به نام ارزش در معرض خطر (VaR) جهت تعیین بیشترین ضرر که ممکن است با احتمال مشخصی و در دوره ی معینی اتفاق بیافتد به کار گرفته شده است. اگرچه ارزش درمعرض خطر ، تحت کنترل یک سری محدودیت قرار دارد، با این حال این معیار توسط کمیته بال (یا بازل) به عنوان روش استاندارد اندازه گیری ریسک در پورتفوی بازار انتخاب شده است. مدلهای زیادی برا ی محاسبه ی ارزش در معرض خطر (VaR) وجود دارد.انتخاب مدل به ماهیت پورتفوی و داده هایی که برای تخمین زدن پارامترها استفاده می شوند بستگی دارد.[۱۰]

در ادامه، به پژوهش های داخلی و خارجی که در زمینه های الگوریتم ژنتیک ، الگوریتم های طبقه بندی ، روش های آن و مفهوم ارزش در معرض خطر، می باشد، اشاره می شود.

۲-۱- الگوریتم ژنتیک و انتخاب سبد بهینه

مطهره مقدسی (۱۳۸۹) در مقاله ای تحت عنوان انتخاب و بهینه سازی سبد سهام با استفاده از الگورتیم ژنتیک براساس تعریف متفاوت از ریسک، ۱۴۶ شرکت را برای مطالعه انتخاب کرده است. با در دست داشتن قیمت ماهانه سهام این شرکت ها، ریسک و بازده ماهانه آن ها به عنوان ورودی های الگوریتم محاسبه شده است. سپس ۲ مدل طراحی شده، که مدل اول با وارد کردن محدودیت های بازار واقعی و اعمال نمودن ترجیحات متفاوت سرمایه گذران بر مدل میانگین – واریاس مارکوتیر، مدل توسعه یافته میانگین – واریانس مارکوتیر است و مدل دوم با در نظر گرفتن نیمه واریانس به جای واریانس طراحی شده است. نتایج نشان دهنده کارایی بالای الگوریتم در اندازه های مختلف سبدها بوده اند و بنا بر فرضیه اول ، بین میانگین بازدهی سرمایه گذاری در سبدهای سهام انتخاب و بهینه شده بر مبنای مدل میانگین – نیمه واریانس و توسط الگورتیم ژنتیک و میانگین بازدهی سرمایه گذاری در سبدهای سهام انتخاب شده برمبنای مدل میانگین – واریانس و توسط الگورتیم ژنتیک تفاوت معنی داری وجود ندارد و در مورد واریانس در فرضیه دوم نیز همین طور. نتایج حاکی از کارایی بالای الگوریتم ژنتیک در حل اینگونه مسائل بوده و به عقیده محقق در

نظر گرفتن محدودیت عدد صحیح به عنوان یکی از محدویت های اضافه شده به مدل سنتی باعث شده است جواب ها واقعی تر و سرمایه گذاران راضی تر باشند.

محمدی استخری (۱۳۸۶)، در مقاله ای تحت عنوان انتخاب یک سبد سهام از بین شرکت های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از مدل بهینه سازی الگوریتم ژنتیک ، به وجود اختلاف معنی دار و برتری قابل توجه نتایج روش الگوریتم ژنتیک برای سبدهای ۱۰، ۲۰ و ۳۰ سهمی نسبت به سبدهای نخست روش تصادفی در مورد ۲ متغیر ریسک و بازده دست یافت است.

سفیانه و بنبوزیانه[۸] (۲۰۱۲) ، در مقاله ای تحت عنوان انتخاب پورتفوی با استفاده از الگوریتم ژنتیک ، الگوریتم ژنتیک را بر روی مثال ساده ای شامل ۵ سهم به کار بسته اند.آنها ذکر می کنند که نتایج بدست آمده جالب توجه بوده و کارایی الگوریتم ژنتیک را برای همگرایی سریعش به سمت راه حل بهتر تأیید می کند و همچنین در مورد زمان محاسباتی نیز جالب توجه است.

رودییر[۹] (۲۰۰۷) در رساله ی ارشد تحت عنوان بهینه سازسی پورتفوی و الگوریتم ژنتیک ، به دنبال دستیابی به روش هایی با قدرت، جهت بهینه سازی پورتفوی می باشد. برای این کار او تابع هدف چند هدفه را طراحی کرده است. در این پژوهش ذکر گردیده که جهت محاسبه ریسک از نیمه واریانس استفاده می شود که هم آسان است و هم آسان به کار بسته می شود. این رساله ۲ دیدگاه متفاوت را برای بهبود انتخاب پورتفوی امتحان می کند:

۱. استفاده از روش ریسک/بازده

۲. استفاده از نمونه گیری مجدد برای کاهش اثر اختلال روی فرآیند بهینه سازی

و معتقد است که ترکیب این روش ها نتیجه ی بهتری را در بر خواهد داشت.

۲-۲ الگوریتم های طبقه بندی

طبقه بندی تشکیل طبقات با استفاده از واحد های (نقاط) جامعه ی مورد بررسی است به طوری که واحدهای درون هر طبقه دارای حداکثر شباهت و واحدهای بین طبقات دارای حداکثر اختلاف باشند.[۴]

یک طبقه بندی مناسب باید منجر به تولید طبقاتی شود که حداکثر تشابه درون طبقات و حداقل تشابه بین طبقات حاصل شود. مبحث طبقه بندی اساس بسیاری از سیستم های پشتیبانی تصمیم را تشکیل می دهد.[۵]

قره چلو (۱۳۸۹)، در مقاله ای تحت عنوان ارزیابی و مقایسه الگوریتم های مختلف طبقه بندی در تهیه نقشه شوری سطحی خاک، به مقایسه روش های مختلف الگوریتم طبقه بندی پرداخته است. در این مقاله روش حداکثر احتمال به عنوان دقیق ترین و روش حداقل فاصله تامیانگین 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 که طبقات از یک توزیع نرمال برخودار باشند.

جیانسی[۱۰] (۲۰۰۷) در مقاله ای تحت عنوان الگوریتم طبقه بندی کارا با استفاده از ماشین بردار پشتیبانی(SVM)[۱۱] برای داده های وسیع، ماشین بردار پشتیبانی رایکی از روش‌های یادگیری بانظارت معرفی کرده است که از آن برای طبقه‌بندی استفاده شده است. به عقیده ی پژوهشگر ابزار مذکور شالوده ی تئوریکی نیرومندی دارد و الگوریتم طبقه بندی بر مبنای آن ، کارایی خوبی را موجب می شود. این ابزار دقت طبقه بندی را افزایش می دهد.

۲-۲-۱- روشهای طبقه بندی

دسته بندی های متفاوتی برای روشهای طبقه بندی در پژوهش ها آمده است.

هاچلوفی[۱۲] (۲۰۱۲) روش های طبقه بندی را به ۲ دسته تقسیم کرده است، روش طبقه بندی خود کار و روش تخصیص. وی این روش ها را اینگونه تعریف می کند: روشهای طبقه بندی خودکار روش هایی هستند بر مبنای فراگیری (یادگیری) نظارت نشده . هدف این روش تشکیل طبقه های همگنی است که با توجه به معیار هدف ایجاد شده اند. روش تخصیص روشی است بر مبنای یادگیری نظارت شده که از مثال ها و شاخص های از قبل شناخته شده بهره می گیرد.

قره چلو (۱۳۸۹) در مقاله ای تحت عنوان ارزیابی و مقایسه الگوریتم های مختلف طبقه بندی در تهیه ی نقشه ی شوری سطحی خاک ، روشها ی طبقه بندی را به دو دسته ی نظارت شده و نظارت نشده تقسیم کرده است. در این مقاله اینطور بیان می شود که روش طبقه بندی نظارت نشده اصولاً بیش شناختی را در مورد مشخصات طبقه ها طلب نمی نماید ، بنابراین نتایج حاصل نیاز به استفاده از اطلاعات جنبی ، تکمیلی ، تفسیر و شناسایی دارد. روش های طبقه بندی نظارت شده بر پایه پیش شناخت دقیق طبقه ها و معرفی آنها بر اساس شناخت قبلی استوارند. در این مقاله ذکر می شود که جهت طبقه بندی نظارت شده ، الگوریتم یا طبقه بندی کننده [۱۳]های مختلفی وجود دارد که در تحقیق بسته به هدف ، وضعیت کلاس ها و دقت مورد نظر از الگوریتم های گوناگونی استفاده می شود که از جمله آن می توان به روش جعبه ای ،.حداقل فاصله تا میانگین و حداکثر احتمال اشاره نمود.

رضایی در پژوهشی تحت عنوان روش های مدرن طبقه بندی ، الگوریتم های فرازی(تقسیمی) ، الگوریتم های سلسله مراتی ، مبتنی بر چگالی ، مبتنی بر مدل و روش های فازی را به عنوان روش های طبقه بندی نام برده است.

۲-۳- ارزش در معرض خطر VaR :

VaR ریشه در بحران‌های مالی دارد که در اوایل دهه ۱۹۹۰ گریبان‌گیر شرکت‌های اورنج کانتی، بارینگز، متال‌گسل‌شفت، دایوا و دیگر شرکت‌ها شد. فصل مشترک این بحران‌ها، این است که ممکن است میلیاردها دلار به علت ضعف در نظارت و مدیریت مالی از دست برود. این مشکل سبب اقبال مؤسسات مالی و ناظران بانک‌ها به VaR ، به عنوان روشی قابل فهم برای به کمیت در آوردن ریسک بازار شد.

VaR بیشترین زیان مورد انتظار در یک بازه زمانی مشخص و در سطح اطمینان معین را اندازه‌گیری می‌کند. سهام‌داران و مدیران مؤسسات مالی می‌توانند تصمیم‌گیری کنند که آیا با این سطح از ریسک، می‌توانند آسوده خاطر باشند یا خیر. اگر پاسخ منفی باشد، باید پروسه محاسبه VaRدوباره طی شود تا تعیین شود که ریسک در کجا باید اصلاح شود.

VaR برخلاف سنجه‌های سنتی ریسک، نمایی‌کلی و جامع از ریسک پرتفوی،که برای محاسبه میزان بدهی به دارایی و همبستگی‌ها و وضعیت‌های جاری به کار می‌رود، ارائه می‌کند . در نتیجه VaR، سنجشی واقعی از ریسک با نگاهی آینده‌نگر است.

جیمز و گروتکه[۱۴] (۲۰۱۰) در مقاله ای تحت عنوان معرفی ریاضیات ریسک در معرض خطر(VaR) ، به دنبال ریاضیات بکار گرفته شده پشت این مفهوم بنیادی هستند، در این مقاله خصوصاً در مورد تئوری احتمالی مربوط، که در پشت مفهوم VaR قرار گرفته بحث می شود، همچنین چند مفهوم از جبر خطی و آن دسته از ریاضیات کاربردی که در محاسبه ی VaR مفید هستند. در این مقاله یک سری محاسبات ریاضی مفید برای محاسبه VaR معرفی می شود.

۲-۴- روش k-means

روش k میانگین (K-means) یکی از روش های مبتنی بر فراز (تقسیمی) است. این الگوریتم در مواردی دارای کاربرد می باشد که در آن هر عارضه تنها به یک کلاس تعلق می گیرد .این الگوریتم یک الگوریتم نظارت نشده و دارای تکرار می باشد که در آن مجموعه داده ها به k طبقه تقسیم بندی شده ونقاط داده به طور تصادفی به این طبقه ها تعلق می گیرند. پس برای هر نقطه ، فاصله ی آن نقطه تا مرکز طبقه محاسبه گردیده و نقطه مورد نظر به نزدیک ترین طبقه تعلق می گیرد. این مراحل تا جایی تکرار می یابد که دیگر هیچ نقطه ای تغییر مکان ننماید.

تلگارس کی و واتانی[۱۵] (۲۰۱۰) در مقاله ای تحت عنوان متود هارتیجان : دسته بندی K-means ، بیان می دارد که روش هارتیجان برای دسته بندی K- means روشی ابتکاری و به شرح ذیل است: انتخاب کردن نقطه و به صورت بهینه ای علامت گذاری آن . این مقاله دو روش دیگر

ابتکاری را نشان می دهد؛ یکی هدایت کردن به سمت تعدادی از دارائی های با ثبات (صفات سازگار) است و دیگری نشان دادن اینکه تفکیک داده ها کاملاً از تفکیک ورونوی[۱۶] غالب ، جدا است.

۳- فرضیات پژوهش

فرضیه های پژوهش به صورت زیر بیان می گردد:

فرضیه اول: ریسک بدست آمده با استفاده از الگوریتم طبقه بندی و الگوریتم ژنتیک، در پورتفوی سهام حاصل، کمتر از ریسک حاصل از بکارگیری روش الگوریتم ژنتیک به تنهایی است.

فرضیه دوم: بازده بدست آمده توسط الگوریتم طبقه بندی و الگوریتم ژنتیک، در پورتفوی سهام حاصل، بیشتر از بازده حاصل از بکارگیری الگوریتم ژنتیک به تنهایی است.

فرضیه سوم: پورتفوی بدست آمده با استفاده از الگوریتم ژنتیک در دسته ای جای می گیرد که با استفاده از روش k-means آن را بهینه تر یافتیم.

۴- روش شناسی پژوهش

جامعه ی آماری این تحقیق کلیه ی شرکت های پذیرفته شده در سازمان بورس اوراق بهادار تهران می باشند. با توجه به اینکه در جامعه ی آماری مورد نظر شرکت ها و صنعت های مختلف جای گرفته اند، می توان جامعه را به طبقه های مختلف تقسیم کرد به نحوی که هر طبقه دارای صنایع همگنی باشد. نمونه با اعمال یک سری محدویت ها از میان جامعه انتخاب شد:

- کلیه ی شرکت هایی که تا پایان سال ۱۳۸۷ در بورس اوراق بهادار تهران پذیرفته شده و تا پایان سال ۱۳۹۰ همچنان در فهرست بورس، باقی مانده باشند.
- اطلاعات مالی ۳ ساله آنها موجود باشد و دسترسی به آن امکان پذیر باشد.

با توجه به این معیارها و همچنین توجه به شرکت هایی که در محدوده ی این ۳سال فعال تر بوده اند و نیز توجه به کلیه ی صنایع، ۱۰۰ شرکت از میان جامعه ی مورد بررسی، به عنوان نمونه برگزیده شد.

برای آگاهی از چگونگی توزیع داده هاابتدا چولگی و کشیدگی داده ها آزمون می شود. چولگی معیاری از تقارن یا عدم تقارن تابع توزیع می‌باشد. نتایج حاصل از این آزمون برای داده های این پژوهش به صورت جدول زیر می باشد:

جدول ۱: نتایج حاصل از آزمون چولگی و کشیدگی

Descriptive Statistics

	N	Mean	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
میانگین بازده	100	.000674	1.200	.241	4.562	.478
ارزش در معرض خطر	100	4.630055E11	5.426	.241	30.655	.478
Valid N (listwise)	100					

با توجه به جدول از لحاظ کجی متغیر بازده نرمال بوده و توزیع آن متقارن است، اما متغیر ریسک نرمال نبوده و دارای چولگی مثبت است. از لحاظ کشیدگی نیز متغیر ریسک و بازده نرمال نبوده و دارای کشیدگی مثبت زیادی می باشند.

پس از بررسی عادی یا نرمال بودن کشیدگی و چولگی توزیع داده‌ها، از آزمون شاپیرو-ویلک و آزمون کولموگروف-اسمیرنوف استفاده می شود تا از نرمال بودن داده‌ها اطمینان حاصل گردد. نتایج این آزمون نیز در جدول زیر نمایش داده شده:

جدول۲: نتایج حاصل از آزمون کولموگروف-اسمیرنوف و شاپیرو-ویلک

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
میانگین بازده	.096	100	.023	.932	100	.000
ارزش در معرض خطر	.417	100	.000	.272	100	.000

a. Lilliefors Significance Correction

با توجه به جدول فوق و مقادیر سطح معناداری برای متغیرها می توان گفت که توزیع متغیرها نرمال نمی باشد. حال با رسم نمودارهای جعبه ای ،Q-Q و نمودار شاخه و برگ می توان به طور جزیی به نحوه ی توزیع داده ها پی برده و علت نرمال نبودن داده ها را مشاهده نمود. پس از رسم این نمودارها، با توجه به نتایج حاصل و طبق نظر متخصصان، ۹شرکت که دارای داده هایی پرت بودند از نمونه ی آماری ما حذف شده و بدین ترتیب ۹۱ شرکت به عنوان نمونه ی نهایی انتخاب شدند.

تحقیق حاضر از بعد هدف کاربردی و از منظر داده ها کمی است. اطلاعات مورد نیاز پژوهش، از کتاب ها، مقالات، مجلات فارسی، لاتین و پایان نامه ها می باشد و داده های شرکت ها از بانک

اطلاعاتی ذخیره شده در آرشیو بورس اوراق بهادار تهران حاصل شده است. متغیر های اصلی تحقیق بازده و ارزش در معرض خطر می باشد، که محاسبات برای هرشرکت عضو نمونه، به طریق زیر صورت گرفته است :

الف)بازده روزانه ی شرکت ها، بر اساس تغییرات قیمت روزانه سهام شرکت ها، از طریق فرمول زیر، محاسبه شده است.

$$R_{t+1} = Ln\frac{P_{t+1}}{P_t}$$

R_{t+1}: بازده لگاریتمی امروز ، P_{t+1}: قیمت امروز سهم و P_t: قیمت سهم در روز گذشته

به این ترتیب برای هر شرکت بازده روزانه ی ۳ سال آن محاسبه شده است و نهایتاً میانگین بازده هرشرکت برآورد شده است.

ب)ریسک شرکت که از معیار ارزش در معرض خطر بهره گرفته می شودو از فرمول زیر برای محاسبه ی آن استفاده شده است.

$$VaR = N \cdot Z\alpha \cdot \delta \sqrt{T}$$

N : ارزش بازار دارایی (سهام) ، $Z\alpha$: سطح اطمینان ، δ : انحراف معیار بازده ، T: تعداد روزها

۵- مدل‌ها و الگوریتم های پژوهش

۵-۱ الگوریتم k-means

خوشه بندی k-means یکی از روش های طبقه بندی نظارت نشده و غیر سلسله مراتبی می باشد. K-means برای خوشه بندی داده ها مراحل زیر را طی می کند:

گام اول: تقسیم بندی موارد به k خوشه ی اولیه.

گام دوم: در طول لیست موارد مجموعه ی داده ها، پیش رفته و هر مورد را به طبقه ای که مرکزش(میانگینش) نزدیک ترین است، اختصاص می دهیم. فاصله معمولاً با استفاده از معیار اقلیدسی تخمین زده می شود. برای خوشه ای که مورد جدید را گرفته و همینطور برای خوشه ای که آن مورد را از دست داده مجدداً مرکز طبقه محاسبه می شود.

گام دوم آنقدر تکرار می شود تا هیچ تخصیص مجدد دیگری عملی نباشد.

۵-۲ الگوریتم ژنتیک

الگوریتم های ژنتیک (GA) روشهای جستجو وبهینه سازی احتمالی هستند که بر اساس مفاهیم برگرفته از تکامل بیولوژیکی طبیعی مانند تولید مثل ، جفت گیری و جهش ساخته می

شوند. الگوریتم ژنتیک برای یافتن پورتفوی بهین مراحل زیر را طی می کند. نرم افزار مطلب جهت اجرای الگوریتم به کار برده شده است.

الف) ارزش دهی آغازی(انتخاب جمعیت)

جمعیت، مجموعه ای از کروموزوم هاست که از k ژن ترکیب شده اند. این جمعیت، در ابتدا به صورت تصادفی از کدهای موجود استفاده می کند.

ب) تابع تناسب[۱۷]

این مرحله شامل سنجش کروموزوم های تولید شده از عمل قبل، با تابع تناسب است. طراحی این تابع، نقطه ی بسیار مهمی در به کارگیری الگوریتم ژنتیک (GA) است.[۱۵] تابع برازندگی ای که در این پژوهش استفاده می شود به شکل زیر است:

$$E = W_r \frac{1}{1 + \sum_{i=1}^{N} R_i} + W_v \sum_{i=1}^{N} V_i$$

در این تابع، R نشان دهنده ی بازده و W_r ضریب اهمیت بازده، V ارزش در معرض خطر و W_v ضریب اهمیت این پارامتر می باشد. اگرچه در این پژوهش W_r و W_v یکسان در نظر گرفته شده اما چنان چه بخواهیم برای هرکدام از پارامترها ارزش بیشتری قائل شویم، می توان ضریب اهمیت آن پارامتر را نسبت بیشری در نظر گرفت.

الگوریتم ژنتیک، تابع مذکور را مینییم کرده و بدین صورت به دلیل وجود جمله ی $\frac{1}{1 + \sum_{i=1}^{N} R_i}$ در تابع تناسب، بازده های بیشتر و همچنین وجود جمله ی $\sum_{i=1}^{N} V_i$ در تابع، ریسک های کمتر انتخاب می شود. همانطور که گفته شد وزن سهم ها در هر سبد یکسان در نظر گرفته شده و با توجه به این که سبدهای با اندازه ی یکسان با هم مقایسه شده است از میانگین گیری صرف نظر و از جمع پارامترها در تابع هدف بهره گرفته شده است، که می توان به جای آن از میانگین گیری نیز بهره برد.

ج) عملیات انتخاب

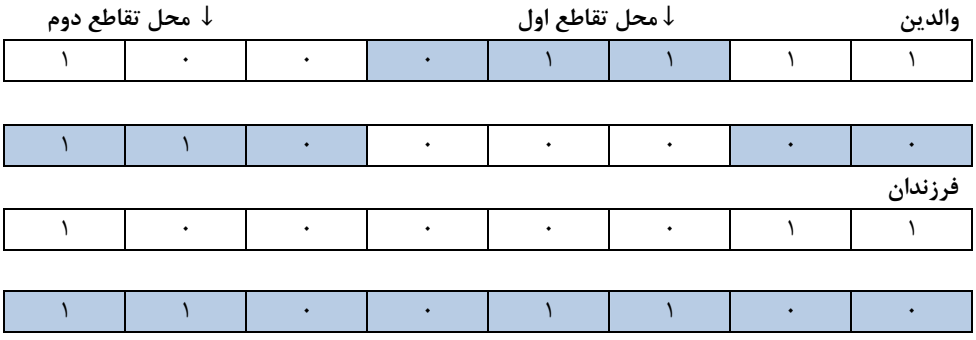
بعد از عمل برآورد(تخمین) جمعیت، بهترین کروموزوم ها(جواب ها) با استفاده از روش انتخاب بر اساس چرخ رولت[۱۸] ، انتخاب می شوند؛ در این روش کروموزومی که عدد برازش(تناسب) بیشتری داشته باشد احتمال بیشتری برای انتخاب شدن دارد.

د) عملیات تقاطع[۱۹]

پس از استفاده از روش انتخاب برای انتخاب دو زوج، عملگر جفت گیری(تقاطع) را برای نقطه ای روی این جفت به کار می گیریم. این عملگر هر والد را به دو بخش با وضعیت مشابه، تقسیم می کند[۱۵]. فرزند اول از قسمت اول والد اول و قسمت دوم والد دوم ساخته شده، در حالی که فرزند دوم از قسمت دوم والد اول و قسمت اول والد دوم ساخته شده است.[۱]

شکل۱. عملیات در ۱ نقطه ی تقاطع

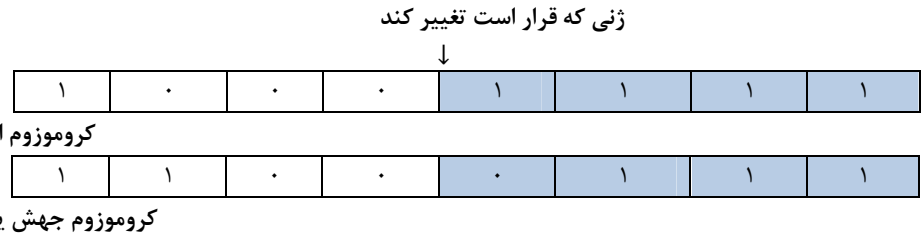
روش بالا تنها یکی از روش های تقاطع یا به عبارتی دیگر بازترکیبی می باشد. روشی که ما از آن بهره برده ایم روش ادغام دو نقطه ای[۲۰] می باشد. این روش مانند روش قبل است با این تفاوت که کروموزوم ها دارای دو محل تقاطع هستند.



شکل۲. عملیات در ۲ نقطه ی تقاطع

ه) عمل جهش[۲۱]

این عمل، قرعه کشی ژن تصادفی جدید در کروموزوم و جابه جا کردن آن با ژن تصادفی رایج است.[۱۵] این عملگر، ویژگی تصادفی بودن و امکان فرار از نقاط محلی را فراهم می آورد.[۱]



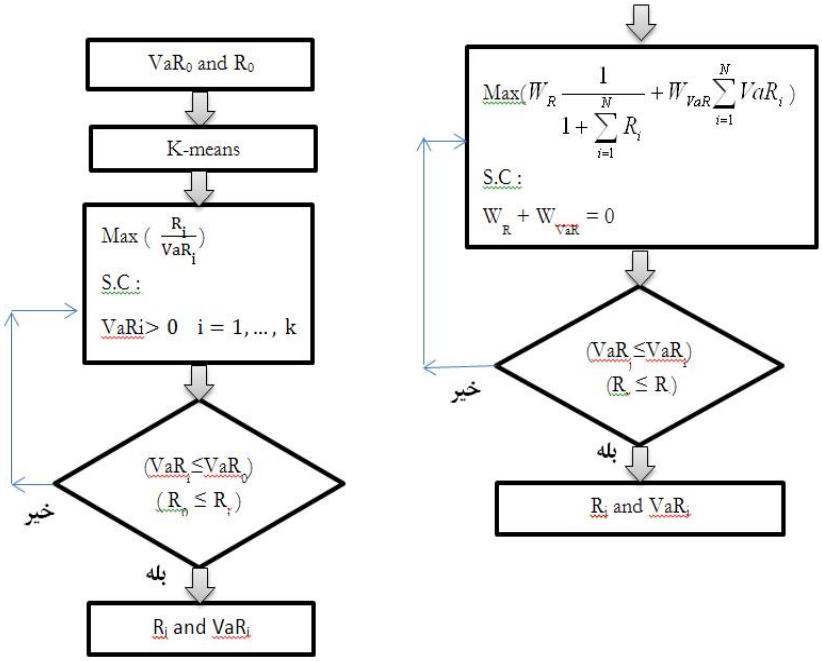
شکل۳. عملیات در نقطه ی تقاطع

ی) شرایط همگرائی[۲۲]

در این مرحله، تولید نهایی در نظر گرفته می شود. اگر نتیجه مطلوب باشد، کروموزوم بهینه حاصل شده است. در غیر این صورت مرحله ی انتخاب و تولید مثل تا رسیدن به نسل های ویژه و همگرایی جمعیت خاص ادامه می یابد.[۱۵]

بنابراین برپایه ی الگوریتم های ذکر شده ، فرآیند اجرا و به کارگیری مدلها به شرح زیر در دست خواهد بود.

هدف از الگوریتم طراحی شده در تحقیق حاضر، انتخاب مجموعه دارایی است، که علاوه بر دستیابی به بیشترین بازده، دارای کمترین ریسک باشد. برای نیل به این هدف، در گام اول با استفاده از روش k-means داده ها را طبقه بندی کرده، به گونه ای که داده ها بیشترین شباهت را نسبت به داده های هم گروه و بیشترین تفاوت را نسبت به داده های گروه های دیگر دارند. این الگوریتم در نرم افزار مطلب پیاده سازی شده و داده ها طبق نظر متخصصان به ۵ گروه دسته بندی می شوند. سپس در مرحله ی بعد بازده و ریسک هر دسته جداگانه محاسبه شده و دسته ای که بیشترین بازده و کمترین ریسک را دارد به عنوان طبقه ی بهینه و خروجی الگوریتم طبقه بندی انتخاب می شود. تابع برازش تعریف شده در این مقاله، دارای پارامتر اندیس شرکت های انتخابی می باشد و در نهایت، خروجی این الگوریتم، اندیس آن دسته از شرکت هایی است که کمترین میزان تابع برازش را دارا باشند. به عبارت دیگر دارای کمترین مقدار مجموع ریسک و بیشترین مقدار مجموع بازده در میان تمامی حالت های ممکن می باشند.

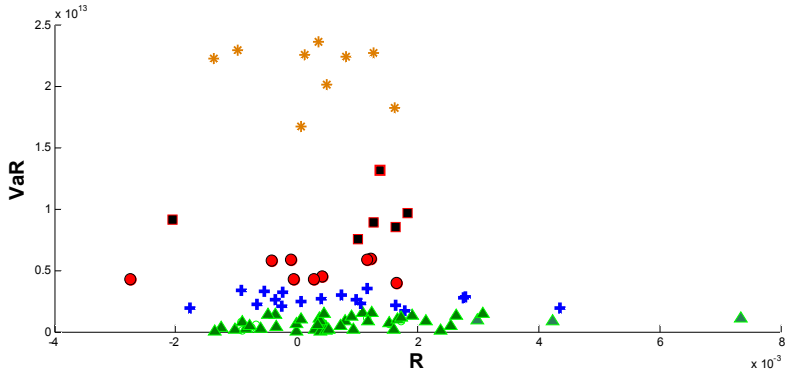


شکل۴: نمودار فرایند تحقیق

۶- یافته هاو نتایج پژوهش

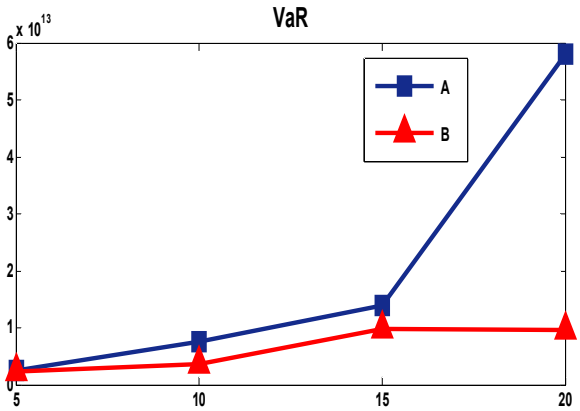
۱۰۰ شرکت به عنوان نمونه انتخاب شد و با انجام آزمون های نرمال و بررسی داده ها ۹ شرکت به دلیل داشتن داده های پرت حذف شد. اطلاعات شرکت ها به صورت روزانه از ۱۳۸۸/۱/۱ تا ۱۳۹۰/۱۲/۳۰ مورد مطالعه و جمع آوری قرار گرفت. پس از محاسبه ی بازده و ارزش در معرض خطر این شرکت ها، الگوریتم طبقه بندی برای دست یابی به طبقه های همگن (طبقه هایی که بازده و ریسک آنها بسیار به هم نزدیک است) انجام شد.

سپس طبقه ای که بیشترین بازده و کمترین ریسک را داشت؛ که شامل ۴۹ شرکت بود، انتخاب شد. شکل ۵ طبقه های به دست آمده از این مرحله را نشان می دهد. محور افقی و عمودی به ترتیب نشان دهنده ی بازده و ارزش در معرض خطر می باشد.

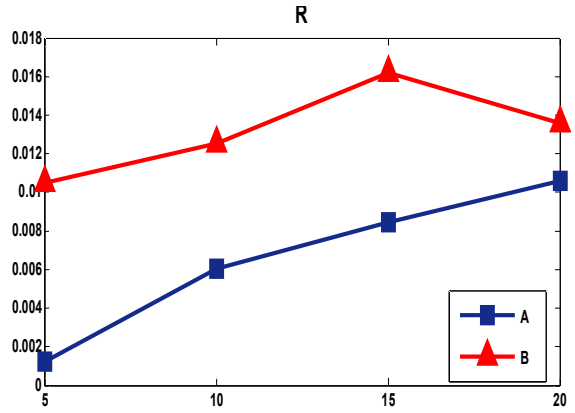


شکل۵ : طبقه بندی با روش k-means

سپس الگوریتم MinVaRMaxR بر روی این طبقه اجرا شده است. با مقایسه ی دو روش برای سبدهای ۵، ۱۰، ۱۵و ۲۰ سهمی، در شکل ۶ مشاهده می شود که بازده بدست آمده با روش ابتدایی یعنی به کارگیری الگوریتم ژنتیک به تنهایی، از روش دوم یعنی اجرای الگوریتم بر روی طبقه ی بهینه، کمتر می باشد.این روش ها را به ترتیب روش A و روش B می نامیم. همچنین با توجه به شکل ۷ مشاهده می شود ریسک حاصل از روش B کمتر از ریسک حاصل از بکارگیری روش A می باشد. محور افقی اندازه ی سبد و محور عمودی مقدار بازده و ریسک را نشان می دهد.

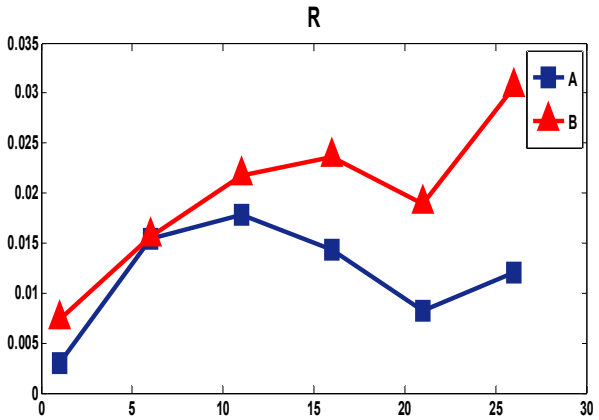


شکل ۶: مقایسه پارامتر بازده در روش A و B

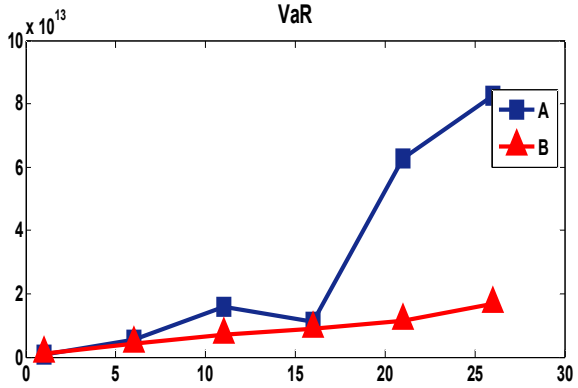


شکل ۷: مقایسه پارامتر ارزش در در معرض خطر در روش A و B

این موضوع برای سبد های بیشتر از ۲۰ سهم نیز صدق می کند. در شکل های ۸ و۹ سبدهای ۲۵ و ۳۰ سهمی نیز قابل مقایسه است.



شکل ۸: مقایسه پارامتر بازده در روش A و B



شکل۹:مقایسه پارامتر ارزش در معرض خطر در روش A و B

این مقایسه نشان دهنده ی کارایی روش B و رسیدن پژوهش به هدف می باشد.

برای اطمینان بیشتر از آزمون ویلکاکسون جهت آزمون فرضیات استفاده شد. زمانی از این آزمون استفاده می شود که از یک جامعه آماری، تنها یک گروه به عنوان نمونه انتخاب و سپس این گروه را از لحاظ یک یا چند متغیر وابسته، قبل و پس از ارائه ی متغیر مستقل مورد بررسی قرار دهیم. این آزمون متناظر با آزمون t است و در صورت وجود نداشتن شرایط آزمون t ، جانشین خوبی برای آن محسوب می شود. آزمون فرض برای هر کدام از ۲فرضیه ی اول بدین گونه می باشد:

$$\begin{cases} \mu_A = \mu_B \\ \mu_A \neq \mu_B \end{cases}$$

جدول زیر نتیجه ی اجراهای مکرر الگوریتم برای سبدهای ۱۰، ۲۰ و ۳۰ سهمی، از ۲روش A و B می باشد.

Ranks

		N	Mean Rank	Sum of Ranks
Return.B – Return.A	Negative Ranks	2[a]	3.00	6.00
	Positive Ranks	13[b]	8.77	114.00
	Ties	0[c]		
	Total	15		
VaR.B – VaR.A	Negative Ranks	14[d]	8.50	119.00
	Positive Ranks	1[e]	1.00	1.00
	Ties	0[f]		
	Total	15		

a. Return.B < Return.A b. Return.B > Return.A c. Return.B = Return.A
d. VaR.B < VaR.A e. VaR.B > VaR.A f. VaR.B = VaR.A

Test Statistics[c]

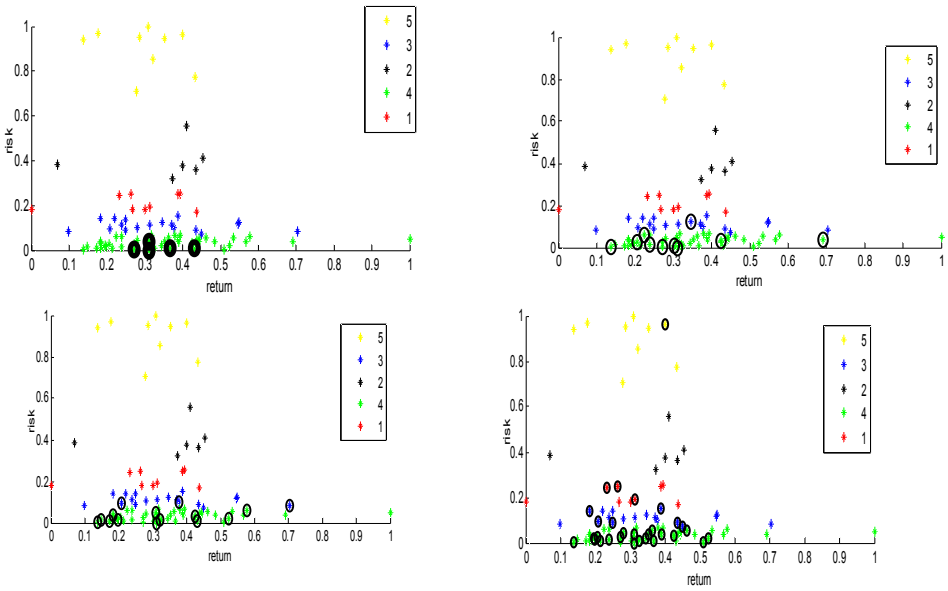
	Return.B – Return.A	VaR.B – VaR.A
Z	-3.067[a]	-3.351[b]
Asymp. Sig. (2-tailed)	.002	.001

a. Based on negative ranks. b. Based on positive ranks. c. Wilcoxon Signed Ranks Test

جدول ۳: نتایج حاصل از آزمون ویلکاکسون

در جدول Rank رتبه ها، میانگین رتبه ها و مجموع رتبه ها بدست آمده است و در جدول Test Statistic معیار تصمیم (.Sig) برای بازده و ریسک به ترتیب مقدار 0.001 و 0.002 بدست آمده است و از آنجایی که این مقدار از 0.05 کوچکتر است لذا فرضیه ی صفر رد شده، یعنی تفاوت معنی داری بین مقدار ریسک و بازده پورتفوی ها، قبل و پس از بکارگیری k-means وجود دارد. بدین معنی که اگر سرمایه گذار هنگام ورود به بازار سرمایه برای انتخاب اینکه چه سبدی از سهام را برگزیند یا به عبارتی کدام یک از شرکت های عضو بورس اوراق بهادار را انتخاب کند تا بازده سرمایه گذاریش افزایش و ریسک سبدش کاهش یابد، ابتدا با استفاده از روش خوشه بندی k-means به طبقه بندی انتخاب ها بپردازد و سپس با انتخاب طبقه ای که بیشترین بازده و کمترین ریسک را دارد و اجرای الگوریتم ژنتیک بر روی آن طبقه ، با طراحی تابع هدفی که به طور همزمان ریسک کمتر و بازده بیشتر را برمی گزیند، در نهایت به نتیجه بهتری (سبد بهینه تری) دست می یابد. این سبد نسبت به سبدی که با استفاده از یک الگوریتم به تنهایی، انتخاب شده است، بهینه تر بوده ، یعنی دارای ریسک کمتر و بازده بیشتر است.

در مورد فرضیه ی سوم، یافته ها نشان دهنده ی این بود که هنگام اجرای ژنتیک بر روی ۹۱ شرکت،در پورتفو های ۵ سهمی و اغلب اوقات ۱۰ سهمی و گاهی ۱۵ سهمی، تمام شرکت ها از میان طبقه ی بهینه ی روش B انتخاب می شوند. اما هرچه اندازه ی پورتفوی بزرگتر می شود، الگوریتم ژنتیک داده هایی را از طبقه های دیگر بر می گزیند که باعث کاهش کارایی آن در مقایسه با روش B و افزایش ریسک می شود. شکل زیر سبدهای ۵، ۱۰، ۱۵ و ۳۰ سهمی را نشان می دهد.



شکل ۹. اجرای الگوریتم ژنتیک برای سبدهای با اندازه ی متفاوت

۷- نتیجه گیری و بحث

نهایتاً این نتیجه حاصل شد که با توجه به پیچیده بودن عمل تصمیم گیری در انتخاب سبد سرمایه گذاری، تلفیق یک سری از الگوریتم ها ممکن است نتایج بهتری را حاصل کند، همانطور که در این پژوهش این نتیجه حاصل شد که اگر با الگوریتم طبقه بندی و سپس انتخاب طبقه ی بهینه برای ورودی الگوریتم ژنتیک، انتخاب ها را برای الگوریتم ژنتیک، محدود کنیم یا به بیانی دیگر به آن دید بدهیم، به هدف خود که افزایش بازده و کاهش ریسک است نزدیک تر می شویم. بنا به یافته ها این روش موجب بهبود اجرای الگوریتم ژنتیک می شود. همچنین وجود انعطلاف در تابع برازش طراحی شده جهت وزن دادن به پارامتر بازده و یا ریسک، موجب می شود که سرمایه گذار، مطابق با اولویت های خویش به پورتفویی بهینه دست یابد. لذا به محققان آتی پیشنهاد می‌گردد که با توجه به اینکه در این پژوهش کلیه شرکت های عضو بورس اوراق بهادار مورد بررسی قرار گرفت ، حال آنکه می توان به بررسی هر صنعت به طور جداگانه پرداخت و شرکت های بهینه هر صنعت، برای قرارگیری در سبد سرمایه گذاری، شناسایی شود.

پیشنهاد دیگر آن است که سایر روش های طبقه بندی مانند FCM ، SOM و غیره نیز جهت دسته بندی داده ها و انتخاب دسته ای برای ورودی الگوریتم ژنتیک ، بررسی و مقایسه شود

فهرست منابع

* البرزی محمود.الگوریتم ژنتیک. چاپ اول. تهران: مؤسسه ی انتشارات علمی دانشگاه صنعتی شریف؛ ۱۳۸۸.
* امیری مقصود، شریعت پناهی مجید و بناکار محمد هادی. انتخاب سبد سهام بهینه با استفاده از تصمیم گیری چند معیاره. فصلنامه بورس اوراق بهادار. ۱۳۸۹. سال سوم. شماره ۱۱.
* بهکیش محمد مهدی. اقتصاد چیست؟. چاپ دوم. تهران: نشر نی؛ ۱۳۸۱.
* حنیفی فرهاد. بحرالعلوم محمد مهدی و جوادی بایک. طراحی و تحلیل مقایسه ای الگوریتم های فرا ابتکاری جهت پیاده سازی سرمایه گذاری شاخص محور در بورس تهران. چشم انداز مدیریت. پاییز ۱۳۸۸. شماره۳۲: ص ۱۰۸-۸۹.
* رضایی علیرضا. روش های مدرن طبقه بندی(مرکز آمار ایران) . دفتر تهیه ی طرح های فنی و چارچوب های آماری.
* سجاد رسول ، گرجی مهسا. برآورد ارزش در معرض خطر با استفاده از روش باز نمونه گیری بوت استرپ(مطالعه موردی بورس اوراق بهادار تهران). فصلنلمه علمی- پژوهشی مطالعات اقتصادی کاربردی در ایران. ۱۳۹۱. سال اول. شماره ۱: صفحات ۱۳۷-۱۶۴.
* شهدایی محمد علی. تحلیل بنیادی در بازار سرمایه ایران. چاپ سوم. تهران: چالش؛ ۱۳۸۶.
* فروزان محمد رضا ، نیرومند محمد رضا. روشهای نوین بهینه سازی. چاپ اول. اصفهان: جهاد دانشگاهی؛ ۱۳۸۸.
* فقیه نظام الدین ، هنرور علی. الگوریتم ژنتیک در برنامه ریزی (بازرسی های پیشگیرانه).چاپ اول. قم: نسیم حیات؛۱۳۸۳.
* محمد استخری نازنین. انتخاب سبد سهام از میان سهام شرکت های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از مدل بهینه سازی الگوریتم ژنتیک. مجله ی توسعه و سرمایه. ۱۳۸۶. شماره ۱: صفحات ۷۱-۹۲.
* مقدسی مطهره. انتخاب سبد سهام با الگوریتم ژنتیک وتعاریف متعددی از ریسک. فصلنامه مدیریت صنعتی دانشگاه علوم انسانی آزاد اسلامی. ۱۳۸۹. سنندج. سال پنجم. شماره ی ۱۱.
* یاری محمدحسن .آکسفورد بورس (اولین دیکشنری بازار بورس اوراق بهادار تهران). چاپ اول. تهران: مؤسسه انتشاراتی مرکز فکر؛ ۱۳۸۶.۹۶۴-۲۶۴۹-۱۷-۹ .

* Bostanci Betul, Bostanci Erkan. An Evaluation of classification algorithms using MC Nemar`s Test.
* Das Nadita. Hedge fund classification using k-means clustering method. 9th International conference on computing in Economics and Finance. 2003. Washington. July 11-13.
* Hachloufi EL. M. Stocks Portfolio Optimization Using Classification and Genetic Algorithm. Mathematical Siences. 2012. Vol 6: 4673-4684.
* Jeyanthi S N. Efficient Classification Algorithm Using SVMs for Large Data set. 2007.
* Jorion Philippe. Value at Risk : The New Benchmark For Managing Financial Risk. 2007. Mc ,Grow-Hill. Third edition.
* Nanda S R , Mahaty B & Tiwari M K. Clustering Indian Stock Market data for portfolio management. Expert systems with Applications. 2010. 37.
* Roudier Felix. portfolio optimization and Genetic Algorithms. zurich. 2007.
* Sefiane Slimane, Benbouziane Mohamed. portfolio selection using genetic algorithm. 2012. Journal of Applied Finance & Banking. vol 2. no4: 143-154.
* White Richard. Methods for classification. 1996 Agu 16

یادداشت‌ها

[1] . Sharpe & Stone
[2] . Rudd & Rosenbeg
[3] . Konno & Yamazaki
[4] . Zenios & Kang
[5] . Speranza
[6] . Yoshimoto
[7] . Hamza & Janssen
[8]. Sefiane & Benbouziane
[9]. Roudier
[10] . Jeyanthi
[11] . Support Vector Machine
[12] . Hachloufi
[13]. Classifier
[14] . James & Grotke
[15]. Telgarsky & Vattani
[16] . Voronoi
[17] . Fitness function
[18] . Roulette Wheel selection
[19] . Operations crossing
[20] . Two-point CrossOver
[21] . Mutation
[22] . Conditions for Convergence