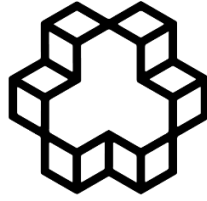


إلى خير



دانشگاه صنعتی خواجه نصیرالدین طوسی

بهینه‌سازی سبد سرمایه در بازار رمزارزها با استفاده از یادگیری ماشین و روش‌های خوشه‌بندی

علیرضا نژادشمسی

استاد راهنما

دکتر امیرعباس نجفی

استاد مشاور

دکتر مجید میرزایی

پایان‌نامه‌ی کارشناسی ارشد

رشته‌ی مهندسی صنایع گرایش مهندسی مالی

تیر ماه ۱۴۰۳

چکیده

همواره یکی از دغدغه‌های بزرگ سرمایه‌گذاران در بازارهای مالی، انتخاب سبدی متشکل از دارایی‌های مختلف موجود در آن بازار بوده است. این مسئله به خصوص در بازار نوظهوری مانند بازار رمزارزها که در معرض انواع ریسک‌های سیستماتیک و غیرسیستماتیک قرار دارد، بیش از دیگر بازارها به چشم می‌آید. انتخاب یک سبد سرمایه‌گذاری بهینه موجب کاهش ریسک سرمایه‌گذاری در عین دریافت بازدهی مناسب می‌شود. این امر موجب می‌شود که سرمایه‌گذاران سنتی که اخیراً به بازار رمزارزها علاقه‌مند شده‌اند و شناخت کمی نسبت به آن دارند، بتوانند نسبت به سرمایه‌گذاری کم‌ریسک‌تر در این بازار اقدام کنند.

پژوهش حاضر با هدف کمک به این دسته از سرمایه‌گذاران، یک مدل چندمرحله‌ای را پیشنهاد می‌کند که در ابتدا نمادهای موجود در بازار رمزارزها را به وسیله‌ی مدل‌هایی مانند انتشار وابستگی، K-Means و K-Medoids خوشه‌بندی می‌کند، سپس در هر خوشه نماد یا نمادهایی را انتخاب می‌کند، و در نهایت نسبت به بهینه‌سازی سبد سرمایه‌گذاری با استفاده از نمادهای باقی‌مانده و به وسیله‌ی روش‌هایی مانند روش هم‌وزن، میانگین-واریانس، میانگین-CVaR، HRP و Sparse اقدام می‌کند.

این پژوهش در نهایت با استفاده از سلسله‌نمودارهای مقایسه‌ای، روش‌های مختلف را با یکدیگر مقایسه می‌کند تا سرمایه‌گذار بتواند با توجه به نیازهای خود بهترین روش را انتخاب نماید. نتایج نشان می‌دهد که بهترین عملکرد از نظر شاخص نسبت شارپ، مدلی است که در مرحله‌ی خوشه‌بندی از روش K-Means، در مرحله انتخاب نماد در هر خوشه از روش بیشترین نسبت شارپ، و در مرحله‌ی انتخاب سبد سرمایه‌گذاری از روش HRP استفاده می‌کند.

واژگان کلیدی: انتخاب سبد سرمایه‌گذاری، بهینه‌سازی پرتفوی، بازار رمزارز، ارزش‌های دیجیتال، یادگیری

ماشین، خوشه‌بندی

فهرست مطالب

فصل اول: مقدمه و کلیات تحقیق..... ۱

۱-۱- مقدمه ۲

۲-۱- هدف از پژوهش ۲

۳-۱- توضیح موضوع پژوهش..... ۳

۴-۱- توجیه، انگیزه و علت انتخاب موضوع..... ۳

۵-۱- اهمیت موضوع ۴

۶-۱- مرور کلی بر ادبیات موضوع..... ۵

۷-۱- جنبه‌های جدید بودن موضوع..... ۶

۸-۱- کاربردهای موضوع پژوهش..... ۷

۸-۱- کاربران نتایج پژوهش..... ۷

۹-۱- جمع‌بندی ۷

فصل دوم: مبانی نظری و مرور ادبیات..... ۸

۱-۲- مقدمه ۹

۲-۲- انتخاب سبد سرمایه‌گذاری..... ۹

۱-۲-۲- مدل هم‌وزن..... ۱۰

- ۲-۲-۲- مدل انتخاب سبد بازار ۱۰
- ۳-۲- بهینه‌سازی سبد سرمایه‌گذاری ۱۱
- ۱-۳-۲- مدل مارکوویتز ۱۲
- ۲-۳-۲- انتخاب نقطه‌ی بهینه روی مرز کارا ۱۳
- ۱-۲-۳-۲- روش کمینه کردن واریانس ۱۴
- ۲-۲-۳-۲- روش بیشینه کردن بازده ۱۵
- ۳-۲-۳-۲- روش بیشینه کردن نسبت شارپ ۱۵
- ۴-۲-۳-۲- روش بیشینه کردن حداقل بازده مورد انتظار ۱۶
- ۵-۲-۳-۲- روش بیشینه کردن مطلوبیت ۱۶
- ۳-۳-۲- توسعه‌ی مدل مارکوویتز ۱۶
- ۱-۳-۳-۲- تغییر سنجه‌ی ریسک ۱۷
- ۲-۳-۳-۲- رویکرد فازی ۱۸
- ۳-۳-۳-۲- رویکرد بهینه‌سازی استوار ۱۸
- ۴-۳-۲- مدل‌های متعادل‌سازی ریسک ۱۹
- ۴-۳-۲- مدل‌های تنک یا Sparse ۱۹
- ۴-۲- یادگیری ماشین ۲۰
- ۱-۴-۲- یادگیری نظارت‌شده ۲۰
- ۲-۴-۲- یادگیری بدون نظارت ۲۱
- ۳-۴-۲- روش‌های رگرسیون ۲۱
- ۱-۳-۴-۲- روش رگرسیون خطی ۲۱
- ۲-۳-۴-۲- روش رگرسیون بردار پشتیبان ۲۲
- ۳-۳-۴-۲- روش رگرسیون چندجمله‌ای ۲۲
- ۴-۳-۴-۲- روش رگرسیون درخت تصمیم ۲۳
- ۴-۳-۴-۲- روش رگرسیون جنگل تصادفی ۲۳
- ۴-۴-۲- روش‌های دسته‌بندی ۲۴

۲۴ روش رگرسیون لجستیک
۲۴ روش درخت تصمیم
۲۴ روش جنگل تصادفی
۲۵ روش بیز ساده
۲۵ روش‌های خوشه‌بندی
۲۵ روش ماشین بردار پشتیبانی
۲۶ روش انتشار وابستگی
۲۶ روش K-Means
۲۶ روش K-Medoids
۲۷ روش خوشه‌بندی فضایی مبتنی بر چگالی در کاربردهای دارای نویز
۲۸ روش خوشه‌بندی سلسله‌مراتبی
۲۸ ۵-۲ مروری بر ادبیات موضوع
۳۴ ۶-۲ جمع‌بندی
۳۵ فصل سوم: روش‌شناسی تحقیق
۳۶ ۱-۳ مقدمه
۳۶ ۲-۳ روش تحقیق
۳۸ ۳-۳ استخراج داده‌های اولیه و تاریخی
۳۹ ۴-۳ فلیتر کردن داده‌ها بر اساس تاریخ
۳۹ ۵-۳ خوشه‌بندی
۴۰ ۱-۵-۳ خوشه‌بندی به روش انتشار وابستگی
۴۰ ۲-۵-۳ خوشه‌بندی به روش K-Means

۴۱K-Medoids خوشه‌بندی به روش
۴۲انتخاب نماد در هر خوشه
۴۳انتخاب سبد
۴۳انتخاب سبد به روش هم‌وزن
۴۳انتخاب سبد به روش میانگین-واریانس
۴۴انتخاب سبد به روش میانگین-CVaR
۴۵انتخاب سبد به روش متعادل‌سازی ریسک سلسله‌مراتبی یا HRP
۴۵انتخاب سبد به روش تنک یا Sparse
۴۶خروجی گرفتن نتایج
۴۷جمع‌بندی
۴۸فصل چهارم: پیاده‌سازی و نتایج پژوهش
۴۹۱-۴ مقدمه
۴۹۲-۴ داده‌های ورودی
۵۱۳-۴ پیاده‌سازی مدل
۵۳۴-۴ نتایج مدل
۵۸۵-۴ تحلیل نتایج
۵۸۱-۵-۴ تحلیل نتایج به تفکیک روش خوشه‌بندی
۶۳۲-۵-۴ تحلیل نتایج به تفکیک روش انتخاب سبد سرمایه‌گذاری
۷۰۳-۵-۴ تحلیل نتایج حذف یا عدم حذف نقاط مغلوب
۷۳۶-۴ جمع‌بندی

فصل پنجم: نتیجه‌گیری و پیشنهادات برای تحقیقات آتی.....۷۴

۵-۱- مقدمه۷۵

۵-۲- جمع‌بندی و خلاصه‌ای از پژوهش۷۵

۵-۳- پیشنهادات برای تحقیقات آتی۷۷

فهرست مراجع۷۹

فهرست شکل‌ها

- شکل ۱-۲: مرز کارا برای مجموعه‌ای از دارایی‌ها در نمودار بازده-ریسک ۱۴
- شکل ۱-۳: چارچوب روش تحقیق ۳۷
- شکل ۲-۳: مثالی از الگوریتم K-Means ۴۱
- شکل ۳-۳: مثالی از الگوریتم K-Medoids ۴۱
- شکل ۱-۴: نمونه‌ای از داده‌های ورودی در مرحله‌ی نخست ۵۰
- شکل ۲-۴: نمونه‌ای از داده‌های ورودی در مرحله‌ی دوم ۵۰
- شکل ۳-۴: نمونه‌ای از خوشه‌بندی نمادها ۵۲
- شکل ۴-۴: نمونه‌ای از ارزیابی عملکرد سبدهای سرمایه‌گذاری ۵۳
- شکل ۵-۴: نتایج خوشه‌بندی انتشار وابستگی با حذف نقاط مغلوب ۶۰
- شکل ۶-۴: نتایج خوشه‌بندی انتشار وابستگی بدون حذف نقاط مغلوب ۶۰
- شکل ۷-۴: نتایج خوشه‌بندی K-Means با حذف نقاط مغلوب ۶۱
- شکل ۸-۴: نتایج خوشه‌بندی K-Means بدون حذف نقاط مغلوب ۶۱
- شکل ۹-۴: نتایج خوشه‌بندی K-Medoids با حذف نقاط مغلوب ۶۲
- شکل ۱۰-۴: نتایج خوشه‌بندی K-Medoids بدون حذف نقاط مغلوب ۶۲
- شکل ۱۱-۴: نتایج روش هم‌وزن با حذف نقاط مغلوب ۶۵
- شکل ۱۲-۴: نتایج روش هم‌وزن بدون حذف نقاط مغلوب ۶۵
- شکل ۱۳-۴: نتایج روش میانگین-واریانس با حذف نقاط مغلوب ۶۶
- شکل ۱۴-۴: نتایج روش میانگین-واریانس بدون حذف نقاط مغلوب ۶۶
- شکل ۱۵-۴: نتایج روش میانگین-CVaR با حذف نقاط مغلوب ۶۷

- شکل ۴-۱۶: نتایج روش میانگین CVaR بدون حذف نقاط مغلوب ۶۷
- شکل ۴-۱۷: نتایج روش HRP با حذف نقاط مغلوب ۶۸
- شکل ۴-۱۸: نتایج روش HRP بدون حذف نقاط مغلوب ۶۸
- شکل ۴-۱۹: نتایج روش Sparse با حذف نقاط مغلوب ۶۹
- شکل ۴-۲۰: نتایج روش Sparse بدون حذف نقاط مغلوب ۶۹
- شکل ۴-۲۱: نتایج مربوط به حذف نقاط مغلوب ۷۰
- شکل ۴-۲۲: نتایج مربوط به نگهداری همه‌ی نمادها به تفکیک روش خوشه‌بندی ۷۲
- شکل ۴-۲۳: نتایج مربوط به نگهداری همه‌ی نمادها به تفکیک روش انتخاب سبد سرمایه‌گذاری ۷۲

فهرست جدول‌ها

جدول ۱-۲: مقایسه‌ی مقالات ۳۲

جدول ۱-۴: نتایج مدل ۵۴

فصل اول

مقدمه و کلیات تحقیق

۱-۱- مقدمه

رمزارز به ارزهای دیجیتالی گفته می‌شود که در آن‌ها از تکنولوژی رمزنگاری استفاده شده و معمولاً به صورت نامتمرکز و بدون دخالت یک سازمان مرکزی عمل می‌کند. پس از آغاز به کار رمزارز بیت‌کوین در سال ۲۰۰۹ میلادی، علاقه‌مندان به فناوری‌های جدید به بررسی و استفاده از آن روی آوردند. مزیت‌ها و پتانسیل‌های موجود در این ارز نوظهور موجب شد تا ارزهای دیجیتال متعددی با کاربردهای جدید متولد شوند و توجه سرمایه‌گذاران در سرتاسر جهان را به خود جلب کند. با ایجاد بازار نوظهور رمزارزها، سرمایه‌گذاران باید سیاست‌های مناسبی را اتخاذ کنند تا بتوانند بازدهی بیشتری در سرمایه‌گذاری داشته باشند. در این پژوهش به بهینه‌سازی سبد دارایی برای افزایش بازده و کاهش ریسک در بازار رمزارزها، و با استفاده از روش‌های یادگیری ماشین پرداخته می‌شود.

۱-۲- هدف از پژوهش

یکی از مهم‌ترین مسائل در سرمایه‌گذاری، انتخاب یک پرتفوی مناسب از دارایی‌های مورد نیاز برای کسب بیشترین بازده با متحمل شدن کمترین میزان ریسک است. سرمایه‌گذاران بسته به میزان ریسک‌پذیری‌شان، می‌توانند از روش‌های مختلفی برای توزیع دارایی‌ها در سبد سرمایه‌گذاری خود استفاده کنند که منجر به کسب بازدهی‌های متفاوتی می‌شود. هدف از این پژوهش، ارائه‌ی روشی مبتنی بر یادگیری ماشین برای

بهینه‌سازی سبد سرمایه‌گذاری و یافتن بهترین رویکرد در بازار رمزارزهاست تا به سرمایه‌گذاران این بازار نوظهور کمک کند تا با توجه به شرایط خاص آن، بهترین سبد از ارزش‌های دیجیتال موجود را ایجاد کنند.

۱-۳- توضیح موضوع پژوهش

سرمایه‌گذاران در بازارهای مختلف در معرض انواع ریسک‌های سیستماتیک و غیرسیستماتیک قرار دارند. با توجه به بالا بودن میزان ریسک سیستماتیک در بازار رمزارزها به علت نوظهور بودن آن و موانع قانونی، سرمایه‌گذاران می‌بایست تا حد امکان اقدام به کاهش ریسک‌های غیرسیستماتیک کنند که به وسیله‌ی تنوع‌بخشی^۱ سبد دارایی انجام می‌شود. بنابراین این سرمایه‌گذاران نیاز دارند که پس از تعیین سیاست‌های سرمایه‌گذاری و تجزیه و تحلیل رمزارزها، اقدام به تشکیل یک سبد سرمایه‌گذاری مناسب از آن‌ها به وسیله‌ی وزن‌دهی به هر دارایی کنند. برای این کار روش‌های متعددی وجود دارد. استفاده از روش‌های نوین مانند یادگیری ماشین در فرآیند انتخاب سبد سرمایه‌گذاری می‌تواند به بهبود نتایج کمک کند.

۱-۴- توجیه، انگیزه و علت انتخاب موضوع

رمزارزها نسبت به ارزش‌های رایج مزایای بالقوه و باارزشی دارند که آن‌ها را به دارایی‌های ارزشمندی برای سرمایه‌گذاری تبدیل می‌کند. بیشتر رمزارزها نامتمرکز و فاقد یک ساختار مرکزی هستند. این مسئله به این معناست که برخلاف ارزش‌های رایج که توسط دولت‌ها و بانک‌های مرکزی کنترل می‌شوند، رمزارزها قابل کنترل توسط هیچ سازمان یا نهادی نیستند و سیاست‌گذاری آن‌ها تنها در پروتکلی انجام می‌شود که در

^۱ Diversification

ابتدا توسط سازنده یا سازندگان آن تعریف و برنامه‌نویسی می‌شود. بنابراین در رمزارزها بر خلاف ارزهای رایج، امکان دستکاری میزان عرضه‌ی پول توسط یک نهاد خاص وجود ندارد.

مزیت دیگر رمزارزها شفافیت آنهاست که فناوری بلاک‌چین^۲ آن را ممکن کرده است. در این فناوری که در اکثر رمزارزهای موجود در بازار از آن استفاده می‌شود، تراکنش‌های همه‌ی افراد در بلاک‌های متصل به یکدیگر ثبت می‌شود که قابل مشاهده برای همه‌ی افراد است. این میزان از شفافیت در هیچ یک از سیستم‌های مالی کنونی یافت نمی‌شود. همچنین متن‌باز بودن این ارزهای دیجیتال علاوه بر بالا بردن میزان شفافیت آنها، امکان استفاده‌ی افراد به صورت ناشناس از شبکه را فراهم می‌کند.

با توجه به مزیت‌های متعدد رمزارزها، رشد زیادی برای آنها در آینده پیش‌بینی می‌شود و از این رو، افراد زیادی از جمله سرمایه‌گذاران بازارهای سنتی، ارزهای دیجیتال را به ترکیب سبد دارایی‌های خود افزوده اند. بنابراین لازم است بیش از پیش به انتخاب و بهینه‌سازی سبد دارایی در این بازار نوظهور توجه شود و ویژگی‌های خاص آن مد نظر پژوهشگران قرار گیرد.

۱-۵- اهمیت موضوع

رمزارزها در کنار مزایای ذکر شده دارای خطرات بالقوه‌ای هم هستند. با توجه به حذف نقش دولت‌ها در این اکوسیستم، همواره موانع قانونی زیادی در برابر استفاده از این ارزهای نوظهور وجود داشته است. همچنین به دلیل ویژگی ناشناس بودن استفاده‌کنندگان بسیاری از رمزارزها و امکان استفاده برای مقاصد غیرقانونی، چالش‌های زیادی در مسائل قانونی مربوط به دولت‌ها وجود دارد. علاوه بر این، هک شدن یک شبکه‌ی خاص و از دست رفتن دارایی‌های سرمایه‌گذاران از خطرات دیگری است که دارندگان رمزارزها را تهدید

^۲ Blockchain

می‌کند. با توجه به خطرات این بازار و رشدهای سریع و حباب‌گونه، سقوط‌های شدیدی هم پیش روی سرمایه‌گذاران به وجود آمده است که ضرورت داشتن استراتژی‌های معاملاتی و بهینه‌سازی سبد دارایی‌ها را دوجندان می‌کند. بنابراین لازم است علاوه بر شناخت رمزارزها و پیدا کردن پروژه‌های ارزشمند، ترکیب آن‌ها در پرتفوی دارایی‌ها به گونه‌ای چیده شود که کمترین میزان ریسک را متوجه سرمایه‌گذاران کند. این پژوهش تلاش می‌کند با کمک روش‌های یادگیری ماشین همچون خوشه‌بندی و رگرسیون در مراحل مختلف چینش سبد سرمایه‌گذاری، این فرایند را بهبود بخشد.

۱-۶- مرور کلی بر ادبیات موضوع

پس از ایجاد بازار رمزارزها، پژوهشگران در بسیاری از مقالات مرتبط با بهینه‌سازی سبد سرمایه‌گذاری، تأثیر افزودن رمزارزها به سبد دارایی سرمایه‌گذاران بازارهایی مثل سهام، ارز، کالا و... را مورد بررسی قرار دادند. بسیاری از این پژوهشگران مانند بریر^۳ و همکاران (۲۰۱۵) و گسمی^۴ و همکاران (۲۰۱۸) اضافه کردن بیت‌کوین به سبد سرمایه‌گذاران را مورد مطالعه قرار دادند؛ اما تعدادی از آن‌ها مانند دمیرالای^۵ و بایراسی^۶ (۲۰۲۱)، تنها به بیت‌کوین بسنده نکردند و افزودن رمزارزهای دیگر به سبد دارایی‌های سنتی را نیز بررسی کردند.

با افزایش کاربرد رمزارزها و تکنولوژی آن، بازار بزرگی از ارزهای دیجیتال شکل گرفت و پژوهشگران بسیاری مانند امبا^۷ و همکاران (۲۰۱۸)، آقامحمدی و همکاران (۱۳۹۹)، براونیس^۸ و مستل^۹ (۲۰۱۹) و پلاتاناکیس^{۱۰}

³ Briere

⁴ Guesmi

⁵ Demiralay

⁶ Bayracı

⁷ Mba

⁸ Brauneis

⁹ Mestel

¹⁰ Platanakis

و همکاران (۲۰۱۸) روش‌های انتخاب سبد رمزارز بدون حضور سایر دارایی‌ها را مورد بررسی قرار دادند. همچنین پژوهشگرانی مانند لیو^{۱۱} (۲۰۱۹) و اسچلینجر^{۱۲} (۲۰۲۰) نیاز مربوط به مقایسه‌ی روش‌های متنوع انتخاب و بهینه‌سازی سبد سرمایه‌گذاری را برطرف کردند.

۷-۱- جنبه‌های جدید بودن موضوع

بازار رمزارزها دارای قدمت بسیار کمی است. نخستین رمزارز نامتمرکز بیت‌کوین است که فعالیت آن از سال ۲۰۰۹ شروع شده است. در سال ۲۰۱۱ رمزارزهای دیگری مانند لایت‌کوین^{۱۳} پا به عرصه‌ی وجود گذاشتند و پس از آن بود که به مرور بازاری از ارزهای رمزنگاری‌شده شامل اتریوم^{۱۴}، ریپل^{۱۵}، دس^{۱۶} و... ایجاد شد. این بازار در طول یک دهه فعالیت خود، فرازونشیب‌های زیادی داشته است.

پس از هر رشد سریع در قیمت رمزارزها، سرمایه‌گذاران جدیدی به فکر سرمایه‌گذاری در آن‌ها می‌افتادند و به این ترتیب ارزش کل بازار رمزارزها به مرور زمان بالاتر می‌رفت. سقوط‌های ناگهانی و خطرات بالقوه باعث شد سرمایه‌گذاران به فکر مدیریت دارایی‌های خود بیفتند و تنها به خرید و نگهداری یک رمزارز خاص اکتفا نکنند. بنابراین نیاز است پژوهش‌های بیشتری در این حوزه انجام شود و به سرمایه‌گذاران کمک کند تا با توجه به نوظهور بودن این بازار، بتوانند با انتخاب و بهینه‌سازی پرتفوی مناسب علاوه بر کسب بازده به خوبی از ریسک‌های آن نیز در امان بمانند.

¹¹ Liu

¹² Schellinger

¹³ Litecoin

¹⁴ Ethereum

¹⁵ Ripple

¹⁶ Dash

۸-۱- کاربردهای موضوع پژوهش

کابرد اصلی این پژوهش کمک به سرمایه‌گذاران برای تشکیل و بهینه‌سازی یک سبد سرمایه‌گذاری از ارزهای دیجیتال برای افزایش بازده و کاهش ریسک است. این کار به کمک وزن‌دهی به هر رمزارز و تقسیم بودجه‌ی سرمایه‌گذاری بین آن‌ها انجام می‌شود.

۸-۱- کاربران نتایج پژوهش

این پژوهش می‌تواند برای سرمایه‌گذاران رمزارزها که قصد نگهداری بلندمدت یا کوتاه‌مدت آن‌ها را دارند مفید باشد. کاربران دیگر این پژوهش، سرمایه‌گذاران سایر بازارها هستند که قصد ورود به بازار ارزهای دیجیتال را دارند و می‌خواهند با صرف کمترین میزان ریسک، از منافع آن استفاده کنند. نتایج این تحقیق همچنین می‌تواند برای محققان اقتصادی، سیاست‌گذاران بازارهای مالی و پژوهشگران مفید باشد.

۹-۱- جمع‌بندی

بازار رمزارزها یک بازار نوظاست که در سال‌های اخیر به طور گسترده‌ای مورد توجه سرمایه‌گذاران قرار گرفته است. با توجه به ریسک بالای این بازار، ضروری است که سرمایه‌گذاران اقدام به مدیریت ریسک سرمایه‌گذاری خود، به خصوص به وسیله‌ی تنوع‌بخشی و با تشکیل سبد دارایی‌ها کنند.

فصل دوم

مبانی نظری و مرور ادبیات

۲-۱- مقدمه

با گسترش رمزارزها در دهه‌ی سوم قرن بیست و یکم، گروه‌های مختلفی از ارزهای دیجیتال پیدا شد که هر یک کارکردها و ریسک‌های خاص خود را دارد. این امر موجب می‌شود که سرمایه‌گذاران تنها به سرمایه‌گذاری در بزرگ‌ترین ارز دیجیتال، یعنی بیت‌کوین بسنده نکنند و به فکر تشکیل سبدی از رمزارزها باشند تا هم از مزایای گروه‌های مختلف این رمزارزها استفاده کنند و هم از ریسک‌هایشان در امان بمانند. بنابراین این سرمایه‌گذاران باید روش‌های بهینه‌سازی سبد سرمایه‌گذاری را بررسی کنند و با توجه به ویژگی‌های این بازار بهترین آن‌ها را برای تشکیل سبد انتخاب کنند. در این فصل به مطالعه‌ی سبد دارایی‌ها و روش‌های انتخاب و بهینه‌سازی آن پرداخته می‌شود.

۲-۲- انتخاب سبد سرمایه‌گذاری

سرمایه‌گذاران با هدف کسب سود در آینده، بخشی از نقدینگی خود را به سرمایه‌گذاری در دارایی‌های مختلف اختصاص می‌دهند. هر کدام از این دارایی‌ها دارای ریسک‌های خاص خود هستند و بازده آن‌ها در آینده نامشخص است؛ اما هر سرمایه‌گذار، انتظاری از بازده دارایی‌ها دارد و بر این اساس اقدام به خرید یا فروش آن‌ها می‌کند. تجربه‌ی سرمایه‌گذاران نشان می‌دهد که برای کاهش ریسک سرمایه‌گذاری می‌توانند سبدی از دارایی‌های مختلف تشکیل دهند تا در صورت افت یک یا چند دارایی، متحمل ضرر کمتری شوند.

در واقع با این کار می‌توان هم در گروه‌های مختلفی از دارایی‌ها سرمایه‌گذاری کرد و هم ریسک کل سرمایه‌گذاری را کاهش داد.

۲-۲-۱- مدل هموزن

مدل هموزن یا انتخاب ساده^۱ که به آن مدل $1/N$ یا EW^۲ نیز گفته می‌شود، سرمایه‌گذاری برابری را برای تمام دارایی‌ها پیشنهاد می‌کند. این روش نیازی به پیشبینی بازده مورد انتظار یا ماتریس کوواریانس بازده دارایی‌ها ندارد و با کمترین زمان و هزینه قابل دستیابی است (ابونوری و همکاران، ۱۳۹۷). پژوهش‌های زیادی انجام شده است که بررسی می‌کند که آیا مدل هموزن می‌تواند نسبت به سایر روش‌های انتخاب سبد برتری داشته باشد یا خیر؛ تا جایی که در برخی از آن‌ها این روش بر بسیاری از روش‌های بهینه‌سازی سبد سرمایه‌گذاری عملکرد بهتری داشته است (پلاتاناکیس و اورکوهارت^۳، ۲۰۱۹).

۲-۲-۲- مدل انتخاب سبد بازار

این روش پرتفوی بازار را برای سرمایه‌گذاری پیشنهاد می‌کند. برای انجام این کار، پس از انتخاب دارایی‌های مورد نظر به هر کدام از آن‌ها وزنی متناسب با ارزش کل بازار آن می‌دهیم. بنابراین نسبت هر دارایی از کل سرمایه برابر با رابطه‌ی ۱-۲ خواهد بود:

$$\text{وزن دارایی} = \frac{\text{ارزش بازاری دارایی}}{\text{مجموع ارزش بازاری دارایی‌های سبد}} \quad (\text{رابطه‌ی ۱-۲})$$

^۱ Naïve

^۲ Equally Weighted

^۳ Urquhart

این روش هم مانند روش ساده، بیشتر برای مقایسه و سنجش عملکرد سایر روش‌های انتخاب و بهینه‌سازی سبد سرمایه‌گذاری استفاده می‌شود.

۲-۳- بهینه‌سازی سبد سرمایه‌گذاری

مسئله‌ی اصلی در بهینه‌سازی سبد سرمایه‌گذاری، انتخاب بهینه‌ی دارایی‌هایی است که با مقدار مشخصی سرمایه می‌توان آن‌ها را تهیه کرد. مارکوویتز (۱۹۵۲) با ایجاد یک نوآوری قابل توجه، مدلی ارائه کرد که به کمک آن می‌توان ریسک سبد سرمایه‌گذاری را به وسیله‌ی تنوع‌بخشی به حداقل رساند. ویلیام شارپ^۴ با مشاهده‌ی مشکلات محاسباتی در حل مدل مارکوویتز، شاخص جدیدی به نام بتا^۵ را پیشنهاد داد که درجه‌ی حساسیت نرخ بازدهی دارایی به تغییرات شاخص را اندازه‌گیری می‌کند. نتایج تحقیقات شارپ و دو تن از همکارانش منجر به پدید آمدن مدلی به نام مدل قیمت‌گذاری دارایی‌های سرمایه‌ای یا CAPM^۶ شد که پارادایمی در حوزه‌ی سرمایه‌گذاری بود (راعی و تلنگی، ۱۳۸۳).

تئوری‌های مالی مثل مدل CAPM دارای پیشفرض‌هایی مثل فرضیه‌ی بازار کارا و عقلایی بودن سرمایه‌گذاری هستند که در چند دهه‌ی اخیر مورد تردید واقع شده است. بر این اساس، بسیاری از سرمایه‌گذاران و محققان این مدل‌ها را دارای اعتبار کافی نمی‌دانند (عباس‌نژاد، ۱۳۸۰). از این رو مدل‌های گسترده و جدیدی برای بهینه‌سازی سبد سرمایه‌گذاری تهیه شده است.

^۴ William Sharpe

^۵ Beta

^۶ Capital Asset Pricing Model

توزیع بازده دارایی‌هایی مثل سهام همواره مورد تحقیقات بسیاری قرار گرفته است و نتایج به دست آمده حاکی از آن است که این توزیع نرمال نیست. بر این اساس، محققان سنجه‌های ریسک نامطلوب را مطرح کردند که بین نوسان‌های مطلوب و نامطلوب تفاوت قائل می‌شود و تنها نوسان‌های پایین‌تر از بازده انتظاری سرمایه‌گذار را ریسک در نظر می‌گیرد. بنابراین این تئوری‌ها بر اساس رابطه‌ی بازده و ریسک نامطلوب به معیارهای انتخاب سبد بهینه می‌پردازد (استرادا^۷، ۲۰۰۷).

۲-۳-۱- مدل مارکوویتز

نظریه‌ی میانگین-واریانس برای اولین بار توسط هری مارکوویتز برای حل مسئله‌ی انتخاب مجموعه‌ی بهینه‌ی دارایی ارائه شد. مارکوویتز این مسئله را به صورت برنامه‌ریزی کوادراتیک با هدف حداقل کردن واریانس پرتفوی دارایی‌ها که سنجه‌ای برای ریسک آن‌ها محسوب می‌شود، با این شرط که بازده انتظاری بیشتر از یک مقدار ثابت باشد مطرح کرد. محدودیت دیگر این مسئله‌ی بهینه‌سازی این است که مجموع متغیرهای تصمیم مسئله یعنی وزن دارایی‌ها برابر با یک باشد و هیچ یک از این اوزان کوچک‌تر از صفر هم نشود (صباحی و همکاران، ۱۳۹۹).

بر طبق مدلی که مارکوویتز (۱۹۵۹) آن را ارائه کرده است، فرض می‌کنیم که n دارایی از اوراق بهادار داریم و هر کدام از آن‌ها را با $i \in \{1, \dots, n\}$ نمایش می‌دهیم. در این صورت بازده ورقه‌ی i -ام یعنی R_i ، یک متغیر تصادفی خواهد بود که میانگین آن را \bar{R}_i و انحراف معیار آن را σ_i در نظر می‌گیریم. همچنین کوواریانس بین بازدهی دو دارایی با σ_{ij} نمایش داده می‌شود. در این صورت شکل استاندارد مدل میانگین-واریانس به صورت رابطه‌ی ۲-۲ خواهد بود.

⁷ Estrada

$$\min Z = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} \quad (\text{رابطه ی ۲-۲})$$

$$\text{s. t.} \quad \sum_{i=1}^n w_i \bar{R}_i \geq d$$

$$\sum_{i=1}^n w_i = 1$$

$$w_i \geq 0 \quad \forall i \in \{1, \dots, n\}$$

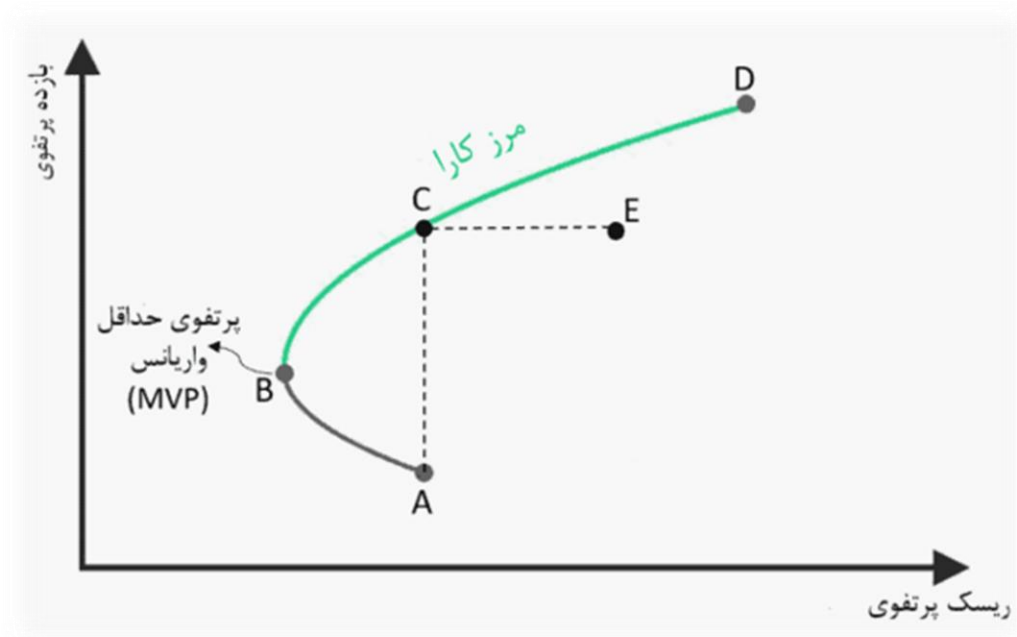
در مدل فوق پارامتر d حداقل بازده مورد انتظار سرمایه‌گذار است و برای مقادیر مختلف آن، جواب‌های متفاوتی به دست می‌آید که مجموعه‌ی آن‌ها، محدوده‌ای به نام مرز کارا را تشکیل می‌دهد. همه‌ی پرتفوی‌های واقع در مرز کارا بهینه محسوب می‌شوند و انتخاب هر کدام یک از آن‌ها بستگی به نظر سرمایه‌گذار و روش‌های ارائه‌شده دارد. (کیانی هرچگانی و همکاران، ۱۳۹۳)

با حل این مسئله‌ی تحقیق در عملیات با به حداقل رساندن تابع هدف، وزن بهینه‌ی هر یک از دارایی‌ها در سبد سرمایه‌گذاری به دست می‌آید. در این صورت برای متغیر تصادفی بازده سبد خواهیم داشت:

$$R_P = \sum_{i=1}^n w_i R_i \rightarrow \begin{cases} E(R_P) = \sum_{i=1}^n w_i \bar{R}_i \\ Var(R_P) = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} \end{cases} \quad (\text{رابطه ی ۳-۲})$$

۲-۳-۲- انتخاب نقطه‌ی بهینه روی مرز کارا

با استفاده از مدل مارکوویتز می‌توان محدوده‌ای شامل پرتفوی‌های بهینه به دست آورد که به آن مرز کارا گفته می‌شود. مرز کارا در نمودار بازده-ریسک مقعر و صعودی است و انتخاب سبد سرمایه‌گذاری روی آن روش‌های مختلفی دارد که در ادامه ذکر می‌شود.



شکل ۱-۲: مرز کارا برای مجموعه‌ای از دارایی‌ها در نمودار بازده-ریسک

۲-۳-۱-۲- روش کمینه کردن واریانس

مارکوویتز (۱۹۵۹) برای انتخاب سبد بهینه روی مرز کارا پیشنهاد داد که حداقل بازده مورد انتظار سهامدار تعیین شود. به این صورت سبدی انتخاب می‌شود که کمترین ریسک را برای بازده تعیین شده فراهم می‌کند. اگر بازده مورد نظر در مرز کارا وجود داشته باشد، نقطه‌ی متناظر با این بازده روی مرز کارا به عنوان پرتفوی بهینه تعیین می‌شود و در صورت پایین‌تر بودن آن از مرز کارا، سبد دارای کمترین ریسک روی مرز کارا (پرتفوی MVP)^۸ جواب مسئله خواهد بود.

^۸ Minimum Variance Portfolio

۲-۲-۳-۲- روش بیشینه کردن بازده

در این روش به جای تعیین کف برای بازده مورد نظر سرمایه‌گذار، از یک سقف برای ریسک مورد انتظار استفاده می‌شود. بنابراین اگر ریسک تعیین شده در مرز کارا موجود باشد، نقطه‌ی متناظر با آن به عنوان سبد بهینه انتخاب می‌شود و اگر این ریسک از مرز کارا بالاتر باشد، پاسخ مسئله بالاترین نقطه روی مرز کارا خواهد بود که دارای بیشترین بازده و ریسک در این محدوده است (کولجک^۹ و همکاران، ۲۰۲۲).

۲-۲-۳-۳- روش بیشینه کردن نسبت شارپ

نسبت شارپ از تقسیم بازده مازاد کسب شده‌ی سبد سرمایه‌گذاری در برابر نرخ بازده بدون ریسک، به انحراف معیار آن به دست می‌آید. روی^{۱۰} (۱۹۵۲) اشاره می‌کند که پرتفویی که دارای بیشترین نسبت شارپ در مرز کاراست، نقطه‌ای است که از نرخ بازده بدون ریسک به آن مماس شود. در واقع در این روش حداقل بازده مورد انتظار به عنوان پارامتر در نظر گرفته می‌شود و برای مماس شدن خط به مرز کارا، شیب آن بیشینه می‌شود. (شهرستانی و همکاران، ۱۳۸۹). نسبت شارپ از رابطه‌ی ۴-۲ به دست می‌آید و هر چه مقدار آن برای یک سبد دارایی بیشتر باشد، نشان‌دهنده‌ی عملکرد بهتر آن است.

$$SR_P = \frac{E(R_p) - R_f}{\sigma_p} \quad (\text{رابطه‌ی ۴-۲})$$

^۹ Čuljak

^{۱۰} Roy

۲-۳-۴- روش بیشینه کردن حداقل بازده مورد انتظار

کاتائوکا^{۱۱} (۱۹۶۳) توضیح می‌دهد که در این روش همانند روش بیشینه کردن نسبت شارپ، خطی مماس به مرز کارا و از نرخ بازده بدون ریسک رسم می‌کنیم؛ با این تفاوت که شیب آن را پارامتری ثابت در نظر می‌گیریم که باید توسط سرمایه‌گذار تعیین شود و حداقل بازده مورد انتظار را بیشینه می‌کنیم (دینگ^{۱۲} و ژانگ^{۱۳}، ۲۰۰۹).

۲-۳-۵- روش بیشینه کردن مطلوبیت

یکی دیگر از روش‌های انتخاب سبد از روی مرز کارا، استفاده توابع مطلوبیت است. به این روش، از طریق منحنی‌های بی‌تفاوتی که برای هر سرمایه‌گذار با توجه به میزان ریسک‌پذیری او متفاوت است، می‌توان نقطه‌ی اشتراک تابع مطلوبیت و مرز کارا را به عنوان پرتفوی بهینه در نظر گرفت. به عنوان یک مثال از توابع مطلوبیت، تابع کوادراتیک رابطه‌ی ۲-۵ قابل تعریف است (لیو، ۲۰۱۹):

$$U = E(R_p) - \frac{\lambda}{2} \sigma_p^2 \quad (\text{رابطه‌ی ۲-۵})$$

۲-۳-۳- توسعه‌ی مدل مارکوویتز

توسعه‌های زیادی برای بهبود عملکرد روش میانگین-واریانس صورت گرفته شده است. در ادامه تعدادی از این روش‌های بهبودیافته ذکر خواهد شد.

¹¹ Kataoka

¹² Ding

¹³ Zhang

۲-۳-۱- تغییر سنجه‌ی ریسک

همان طور که در مدل میانگین-واریانس ذکر شد، مارکوویتز (۱۹۵۲) ریسک دارایی‌ها را با سنجه‌ی واریانس اندازه گرفت. به مرور زمان پژوهشگران سنجه‌های دیگری را برای ریسک در نظر گرفتند و از این طریق مدل مارکوویتز را توسعه دادند. سنجه‌های ریسک به طور کلی به سه دسته تقسیم می‌شود. دسته‌ی اول سنجه‌های مبتنی بر تلاطم است که واریانس نیز یکی از این سنجه‌هاست. شاخص‌های پراکندگی در علم آمار مانند دامنه‌ی تغییرات، دامنه‌ی میان‌چارکی، ضریب تغییرات و... نیز از همین دسته هستند. در این سنجه‌ها نوسان‌های داده‌ها مورد توجه قرار می‌گیرد. دسته‌ی دیگر از سنجه‌های ریسک، سنجه‌های مبتنی بر حساسیت است. این سنجه‌ها حساسیت متغیر تصادفی مورد نظر را در مقابل تغییرات یک متغیر تصادفی دیگر بررسی می‌کنند. از جمله این سنجه‌ها می‌توان به دیرش، تحذب و ضریب بتا اشاره کرد که هر کدام کارکردهای خاص خود را دارا هستند (زمردیان و همکاران، ۱۳۹۸).

آخرین دسته از سنجه‌های ریسک، سنجه‌های ریسک نامطلوب است که بر خلاف سنجه‌های دیگر که تغییرات مثبت و منفی را به عنوان ریسک در نظر می‌گیرد، تنها بخش مربوط به تغییرات منفی و نامطلوب را محاسبه می‌کند. این سنجه‌ها به دو گروه تقسیم می‌شود. نخستین گروه نیم‌سنجه‌ها هستند که شامل نیم‌واریانس، نیم‌بتا و... است (استرادا، ۲۰۰۷)؛ گروه دوم شامل سنجه‌های مبتنی بر صدک مانند ارزش در معرض ریسک، ریزش مورد انتظار و سنجه‌های طیفی است (یامای^{۱۴} و یوشیبا^{۱۵}، ۲۰۰۵).

¹⁴ Yamai

¹⁵ Yoshiba

۲-۳-۳-۲- رویکرد فازی

از دیگر روش‌های توسعه‌ی مدل مارکوویتز که مورد توجه پژوهشگران قرار گرفت، استفاده از رویکردهای دیگر برای عدم قطعیت بود که یکی از این رویکردها، روش فازی است. همان طور که اورتی^{۱۶} و همکاران (۲۰۰۲) نشان داده اند، با توجه به پیش‌بینی‌ناپذیر بودن بازارهای مالی و پیچیدگی آن‌ها ارائه‌ی یک تخمین دقیق از ریسک و بازده مورد انتظار بسیار دشوار است. بنابراین بسیاری از پژوهشگران به استفاده از روش فازی روی آورده اند. در این روش نرخ بازده، بتا، حداقل بازده مورد انتظار و... به جای اعداد قطعی، با اعداد فازی مدل‌سازی می‌شوند. این اعداد فازی می‌توانند به شکل اعداد مثلثی، دوزنقه‌ای یا حالت‌های دیگر باشند. (امیری و محبوب قدسی، ۱۳۹۴)

۲-۳-۳-۳- رویکرد بهینه‌سازی استوار

مدل‌های بهینه‌سازی استوار، بازدهی آینده‌ی دارایی‌ها را به صورت ضرایب غیرقطعی در مسئله‌ی بهینه‌سازی در نظر می‌گیرند و درجه‌ی ریسک‌گریزی سرمایه‌گذاری را به درجه‌ی تحمل در مقابل کل خطای حاصل تخمین بازدهی‌ها تصویر می‌کنند (قره‌خانی و همکاران، ۱۳۹۲). بنابراین همان طور که قهطرانی (۱۳۹۱) اشاره می‌کند، در رویکرد بهینه‌سازی استوار، به دنبال جواب‌های نزدیک به بهینه‌ای هستیم که با احتمال بالا موجه باشند. این رویکرد شامل رویکرد استوار سویستر^{۱۷} (۱۹۷۳)، رویکرد استوار بن تال^{۱۸} و نمیروفسکی^{۱۹} (۲۰۰۰)، و رویکرد استوار برتسیماس^{۲۰} و سیم^{۲۱} (۲۰۰۳) است (پیکانی و روغنیان، ۱۳۹۴).

¹⁶ Orti

¹⁷ Soyster

¹⁸ Ben-Tal

¹⁹ Nemirovski

²⁰ Bertsimas

²¹ Sim

۲-۳-۴- مدل‌های متعادل‌سازی ریسک^{۲۲}

متعادل‌سازی ریسک رویکردی است که برای مدیریت سبد سرمایه‌گذاری، به جای تخصیص سرمایه بر روی تخصیص ریسک متمرکز می‌شود. در این رویکرد انتظار می‌رود که سبد سرمایه‌گذاری ساخته‌شده به واسطه‌ی کاهش ریسک، دارای نسبت شارپ بالاتری نسبت به سایر روش‌ها باشد. این مسئله به خصوص در شرایط افت بازار و وضعیت‌های نوسانی پررنگ‌تر می‌شود. اصل اساسی که سبد تشابه ریسک بر آن استوار است این است که سبد سرمایه‌گذاری باید تا حد امکان بر اساس سهم یکسان ریسک دارایی‌ها از ریسک کل تخصیص یابد. این امر موجب می‌شود که ریسک سبد سرمایه‌گذاری بین دارایی به شکل نسبتاً یکسانی توزیع شود (میرمحمدی و همکاران، ۱۴۰۱).

۲-۳-۴- مدل‌های تنک یا Sparse

روش‌های Sparse در بهینه‌سازی سبد با استفاده از تکنیک‌های انتخاب متغیر، تعداد دارایی‌های موجود در سبد سرمایه‌گذاری را کاهش می‌دهند. این روش‌ها اغلب از تکنیک‌هایی مثل $LASSO^{23}$ استفاده می‌کنند که ضرایب کوچک را به صفر نزدیک می‌کنند و تنها ضرایب مهم را نگه می‌دارند. (تیبشیرانی^{۲۴}، ۱۹۹۶) استفاده از این روش به ما این امکان را می‌دهد که مدل‌هایی با تعداد زیادی دارایی را با انتخاب مجموعه‌ای از دارایی‌های مهم ساده‌سازی کنیم. این کار باعث کاهش پیچیدگی مدل و افزایش قابلیت تفسیر آن می‌شود. همچنین این مدل جلوگیری از بیش‌برازش و بهبود تعمیم‌پذیری آن را در پی دارد. (برودی^{۲۵} و همکاران، ۲۰۰۹)

²² Risk Parity

²³ Least Absolute Shrinkage and Selection Operator

²⁴ Tibshirani

²⁵ Brodie

۲-۴- یادگیری ماشین

یادگیری ماشین یکی از روش‌های تشریح مجموعه‌ای از داده‌ها با استفاده از یافتن یک الگوی جامع است. الگوریتم‌های یادگیری ماشین با استفاده از حجم محدودی از داده‌های گردآوری شده، الگویی برای ارتباط بین داده‌ها پیدا می‌کنند (ندایی و نجفی، ۲۰۱۶). این داده‌های آموزشی برای بهبود عملکرد و ارائه‌ی پیش‌بینی‌های دقیق‌تر استفاده می‌شود. در ادامه به بررسی انواع مسائل یادگیری ماشین پرداخته می‌شود.

۲-۴-۱- یادگیری نظارت‌شده^{۲۶}

مسائل یادگیری ماشین ممکن است نظارت‌شده باشند. در این نوع مسائل علاوه بر داده‌های ورودی، داده‌های خروجی نیز توسط مدل‌ساز به مدل داده می‌شود و مدل موظف به یافتن رابطه‌ای بین ورودی‌ها و خروجی‌هاست. به این شکل مدل می‌تواند با یادگیری داده‌های ورودی و داده‌های خروجی که به آن‌ها داده‌های آموزشی^{۲۷} گفته می‌شود، تابعی برای نحوه‌ی ارتباط این داده‌ها بسازد و از این تابع برای پیش‌بینی داده‌های آتی استفاده کند. یادگیری ماشین نظارت‌شده را می‌توان بر اساس داده‌های خروجی به دو نوع مسئله تقسیم کرد؛ اگر داده‌های خروجی پیوسته باشد به آن مسئله‌ی رگرسیون گفته می‌شود و در غیر این صورت، آن را یک مسئله‌ی دسته‌بندی می‌نامیم (ندایی و نجفی، ۲۰۱۶).

²⁶ Supervised Learning

²⁷ Training data

۲-۴-۲- یادگیری بدون نظارت^{۲۸}

در مدل‌های یادگیری ماشین بدون نظارت بر خلاف مدل‌های نظارت‌شده، داده‌های خروجی به مدل داده نمی‌شود. هدف اصلی در این مسائل، تقسیم‌بندی داده‌ها بر اساس میزان شباهت آن‌هاست. به این شکل که مدل تلاش می‌کند با یادگیری داده‌های ورودی و یافتن ساختاری برای جدا کردن داده‌های مشابه، آن‌ها را به خوشه‌هایی تقسیم‌بندی کند که دارای بیشترین شباهت درون خوشه باشند و از طرف دیگر، کمترین شباهت بین خوشه‌های مختلف وجود داشته باشد. (ساتیا^{۲۹} و آبراهام^{۳۰}، ۲۰۱۳) بنابراین هر یک از خوشه‌ها، گروهی از داده‌های مشابه با هم از نظر ارتباطشان با گروه‌های دیگر است.

تفاوت مسائل دسته‌بندی با این نوع مسائل که به آن‌ها خوشه‌بندی گفته می‌شود، در آن است که «دسته‌ها» توسط مدل‌ساز از پیش تعیین شده هستند و مدل تلاش می‌کند ارتباط داده‌های ورودی با این دسته‌ها را بیابد. این در حالی است که «خوشه‌ها» توسط خود مدل و بر مبنای داشتن بیشترین شباهت درون گروهی و کمترین شباهت بین گروهی ایجاد می‌شوند.

۲-۴-۳- روش‌های رگرسیون

۲-۴-۳-۱- روش رگرسیون خطی^{۳۱}

رگرسیون خطی یک الگوریتم یادگیری ماشینی پرکاربرد برای وظایف رگرسیون است. این یک روش ساده و موثر است که رابطه بین یک متغیر وابسته و یک یا چند متغیر مستقل را مدل می‌کند. الگوریتم با نصب یک خط صاف بر روی نقاط داده‌ای که بهترین نماینده رابطه بین متغیرها هستند، کار می‌کند. این خط

²⁸ Unsupervised Learning

²⁹ Sathya

³⁰ Abraham

³¹ Linear regression

به عنوان خط رگرسیون شناخته می‌شود و برای پیش‌بینی مقدار متغیر وابسته بر اساس مقدار متغیر مستقل استفاده می‌شود. (داتا^{۳۲} و ساهو^{۳۳}، ۲۰۲۱)

۲-۴-۳-۲- روش رگرسیون بردار پشتیبان^{۳۴}

رگرسیون بردار پشتیبان (SVR) یک الگوریتم یادگیری ماشینی است که برای وظایف رگرسیون استفاده می‌شود. این یک روش قدرتمند است که می‌تواند با داده‌های خطی و غیر خطی سروکار داشته باشد. الگوریتم با یافتن یک ابرصفحه که حاشیه بین نقاط داده‌ای و ابرصفحه را بیشینه می‌کند، کار می‌کند. ابرصفحه برای پیش‌بینی مقدار متغیر وابسته بر اساس مقدار متغیر مستقل استفاده می‌شود. (کادام^{۳۵} و همکاران، ۲۰۲۰)

۲-۴-۳-۳- روش رگرسیون چندجمله‌ای^{۳۶}

رگرسیون چندجمله‌ای یک الگوریتم یادگیری ماشینی است که رابطه بین یک متغیر وابسته و یک یا چند متغیر مستقل را مدل می‌کند. این یک نوع رگرسیون خطی است که از توابع چندجمله‌ای برای رسم نقاط داده‌ای استفاده می‌کند. به این شکل که رابطه‌ی بین متغیر مستقل و متغیر وابسته به عنوان یک تابع چندجمله‌ای مدل می‌شود. الگوریتم با یافتن بهترین منحنی مناسب که رابطه بین متغیرها را نشان می‌دهد، کار می‌کند. رگرسیون چند جمله‌ای می‌تواند برای مدل کردن داده‌های غیرخطی استفاده شود. (داتا و ساهو، ۲۰۲۱)

³² Datta

³³ Sahoo

³⁴ Support Vector Regression (SVR)

³⁵ Kadam

³⁶ Polynomial regression

۲-۴-۳-۴- روش رگرسیون درخت تصمیم^{۳۷}

رگرسیون درخت تصمیم یک الگوریتم یادگیری ماشینی است که رابطه بین یک متغیر وابسته و یک یا چند متغیر مستقل را مدل می کند. این یک نوع رگرسیون است که از درخت های تصمیم برای رسم نقاط داده ای استفاده می کند. الگوریتم با تقسیم مجدد داده ها به زیر مجموعه هایی بر اساس مقدار متغیر مستقل کار می کند. سپس زیرمجموعه ها برای ایجاد یک درخت تصمیم استفاده می شوند که رابطه بین متغیرها را نشان می دهد (سارکر^{۳۸}، ۲۰۲۱).

۲-۴-۳-۴- روش رگرسیون جنگل تصادفی^{۳۹}

رگرسیون جنگل تصادفی یک الگوریتم یادگیری ماشینی است که رابطه بین یک متغیر وابسته و یک یا چند متغیر مستقل را مدل می کند. این روش از جنگل تصادفی برای رسم نقاط داده ای استفاده می کند. به این شکل که چندین درخت تصمیم را ترکیب می کند تا دقت پیش بینی ها را بهبود بخشد. الگوریتم با ایجاد چندین درخت تصمیم بر روی زیرمجموعه های مختلف داده ها شروع می شود و سپس نتایج را ترکیب می کند تا پیش بینی نهایی را انجام دهد (لی^{۴۰} و همکاران، ۲۰۲۲).

³⁷ Decision Tree Regression

³⁸ Sarker

³⁹ Random Forest Regression

⁴⁰ Li

۲-۴-۴- روش‌های دسته‌بندی

۲-۴-۴-۱- روش رگرسیون لجستیک^{۴۱}

رگرسیون لجستیک یک الگوریتم ساده و موثر است که رابطه‌ی بین یک یا چند متغیر مستقل و یک متغیر وابسته‌ی دوسویی (به معنای این که مربوط به یک واقعه‌ی تصادفی در دو موقعیت ممکنه است) را مدل می‌کند. الگوریتم با رسم یک منحنی لجستیک بر روی نقاط داده‌ای که بهترین نماینده‌ی رابطه‌ی بین متغیرها هستند، کار می‌کند. این منحنی برای پیش‌بینی احتمال متغیر وابسته برای تعلق به یک کلاس خاص بر اساس مقدار متغیر مستقل استفاده می‌شود. (سارکر، ۲۰۲۱).

۲-۴-۴-۲- روش درخت تصمیم^{۴۲}

درخت تصمیم یک الگوریتم یادگیری ماشینی است که برای مسائل دسته‌بندی استفاده می‌شود. این روش می‌تواند با داده‌های خطی و غیرخطی سروکار داشته باشد. الگوریتم با تقسیم مجدد داده‌ها به زیرمجموعه‌هایی بر اساس مقدار متغیر مستقل کار می‌کند. سپس زیرمجموعه‌ها برای ایجاد یک درخت تصمیم استفاده می‌شوند که رابطه بین متغیرها را نشان می‌دهد (چودھاری^{۴۳} و شوئن^{۴۴}، ۲۰۲۰).

۲-۴-۴-۳- روش جنگل تصادفی^{۴۵}

جنگل تصادفی یک الگوریتم یادگیری ماشینی است که برای وظایف طبقه‌بندی استفاده می‌شود. این یک روش مجموعه است که چندین درخت تصمیم را ترکیب می‌کند تا دقت پیش‌بینی‌ها را بهبود بخشد.

^{۴۱} Logistic Regression

^{۴۲} Decision Tree

^{۴۳} Chowdhury

^{۴۴} Schoen

^{۴۵} Random Forest

الگوریتم با ایجاد چندین درخت تصمیم بر روی زیرمجموعه های مختلف داده ها شروع می شود و سپس نتایج را ترکیب می کند تا پیش بینی نهایی را انجام دهد (لی و همکاران، ۲۰۲۲).

۲-۴-۴-۲- روش بیز ساده^{۴۶}

بیز ساده یک الگوریتم یادگیری ماشینی است که برای وظایف طبقه بندی استفاده می شود. این یک روش احتمالاتی است که رابطه بین متغیر وابسته و یک یا چند متغیر مستقل را مدل می کند. الگوریتم با محاسبه احتمال متغیر وابسته برای تعلق به یک کلاس خاص بر اساس مقدار متغیر مستقل کار می کند (گه^{۴۷} و همکاران، ۲۰۲۰).

۲-۴-۵- روش های خوشه بندی

۲-۴-۵-۱- روش ماشین بردار پشتیبانی^{۴۸}

ماشین بردار پشتیبانی یک روش خوشه بندی است که برای دسته بندی داده های دو دسته ای استفاده می شود. این روش با پیدا کردن ابرصفحه ای که بهترین جداسازی بین داده ها را انجام می دهد، کار می کند. ابرصفحه به گونه ای انتخاب می شود که فاصله بین دو دسته را بیشینه می کند. فاصله به عنوان فاصله ی بین ابرصفحه و نزدیک ترین داده های هر دسته تعریف می شود. SVM یکی از روش های خوشه بندی محبوب است زیرا در فضاهای بالا بعدی موثر است و می تواند با داده های غیرخطی نیز استفاده شود. (ردموند^{۴۹} و همکاران، ۲۰۱۱)

⁴⁶ Naive Bayes

⁴⁷ Goh

⁴⁸ Support Vector Machine (SVM)

⁴⁹ Redmond

۲-۴-۵-۲- روش انتشار وابستگی^{۵۰}

انتشار وابستگی یک روش خوشه‌بندی است که برای دسته‌بندی داده‌های چنددسته‌ای استفاده می‌شود. این روش با پیدا کردن نمونه‌هایی، یا نقاط نماینده، در مجموعه داده که بهترین خلاصه‌ای از داده‌ها را ارائه می‌دهند، کار می‌کند. هر داده بر اساس شباهت آن با نمونه‌ها به یک نمونه نسبت داده می‌شود. شباهت با استفاده از یک اندازه‌گیری از شباهت داده‌ها با همه داده‌های دیگر در مجموعه داده محاسبه می‌شود. انتشار وابستگی یکی از روش‌های خوشه‌بندی محبوب است زیرا سرعت بالا، دقت بالا و تعداد پارامترهای کمتری نسبت به روش‌های دیگر خوشه‌بندی دارد و نیازی به تعیین تعداد خوشه توسط کاربر ندارد (ردموند و همکاران، ۲۰۱۱).

۲-۴-۵-۳- روش K-Means

روش K-Means یکی از روش‌های خوشه‌بندی است که برای دسته‌بندی داده‌های چند بعدی استفاده می‌شود. در این روش، داده‌ها به تعداد k خوشه تقسیم می‌شوند و هر خوشه به گونه‌ای انتخاب می‌شود که مرکز خوشه بهترین نماینده برای داده‌های آن خوشه باشد. این روش برای داده‌هایی که تعداد بعد آن‌ها کم است، به خوبی عمل می‌کند (ردموند و همکاران، ۲۰۱۱).

۲-۴-۵-۴- روش K-Medoids

روش K-Medoids یک تکنیک مؤثر برای خوشه‌بندی داده‌هاست که مشابه با روش K-Means، با انتخاب k نقطه‌ی مرکزی شروع می‌شود، اما برخلاف آن روش، medoid حتماً نقاط واقعی داده هستند. فرآیند

⁵⁰ Affinity Propagation

خوشه‌بندی K-Medoids شامل تکرار مراحل تخصیص هر نقطه به نزدیک‌ترین medoid و سپس به‌روزرسانی medoid با حداقل کردن مجموع فواصل کل تا نقاط خوشه است. این روش به دلیل استفاده از نقاط واقعی داده به عنوان مراکز خوشه، در مقابل داده‌های پرت مقاوم‌تر است و معمولاً نتایج پایدارتری ارائه می‌دهد. (کافمن^{۵۱} و روسو^{۵۲}، ۱۹۸۷).

۲-۴-۵-۵- روش خوشه‌بندی فضایی مبتنی بر چگالی در کاربردهای دارای نویز^{۵۳}

این روش که به اختصار به آن DBSCAN گفته می‌شود، یکی از روش‌های خوشه‌بندی است که برای خوشه‌بندی داده‌هایی با توزیع ناهمگن استفاده می‌شود. در این روش، خوشه‌ها به گونه‌ای تعریف می‌شوند که داده‌هایی که در یک خوشه قرار دارند، از نظر فاصله با یکدیگر نزدیک هستند و داده‌هایی که در خوشه‌ای قرار نمی‌گیرند، از نظر فاصله با داده‌های خوشه‌ها دور هستند. این روش برای داده‌هایی که توزیع آن‌ها ناهمگن است، به خوبی عمل می‌کند (ردموند و همکاران، ۲۰۱۱). این الگوریتم نیازی به تعیین تعداد خوشه توسط کاربر ندارد و خود الگوریتم می‌تواند خوشه‌ها را مبتنی بر غلظت آن‌ها شناسایی کند. این روش یکی از انواع الگوریتم خوشه‌بندی فضایی مبتنی بر چگالی است که بیشتر در موارد مرتبط با کار با داده‌های نویزی استفاده می‌شود.

⁵¹ Kaufmann

⁵² Rousseeuw

⁵³ Density Based Spatial Clustering of Applications with Noise (DBSCAN)

۲-۴-۵-۶- روش خوشه‌بندی سلسله‌مراتبی^{۵۴}

خوشه‌بندی سلسله‌مراتبی یکی از روش‌های خوشه‌بندی داده‌های چندبعدی است. در این روش، داده‌ها به صورت یک درخت خوشه‌بندی شده و هر خوشه به گونه‌ای انتخاب می‌شود که فاصله بین داده‌های آن خوشه کمینه شود. این روش برای داده‌هایی که تعداد بعد آن‌ها کم است، به خوبی عمل می‌کند (ردموند و همکاران، ۲۰۱۱).

۲-۵- مروری بر ادبیات موضوع

با پیشرفت بازار ارزهای دیجیتال، تحقیقات گسترده‌ای در زمینه‌های مختلف روی رمزارزها انجام شده است. از جمله موضوعات این پژوهش‌ها می‌توان به بررسی مسائل مربوط به تکنولوژی، ماهیت، قانون‌گذاری و سرمایه‌گذاری رمزارزها اشاره کرد. یکی از مهم‌ترین دغدغه‌های سرمایه‌گذاران، تشکیل و انتخاب سبد سرمایه‌گذاری است. این مسئله به خصوص از این جهت اهمیت دارد که هنوز صندوق‌های سرمایه‌گذاری بسیار کمی برای رمزارزها وجود دارد و فرد علاقه‌مند به این بازار می‌بایست خود اقدام به بررسی و انتخاب سبد سرمایه‌گذاری کند. از این جهت مقالات مرتبط با تشکیل و بهینه‌سازی سبد سرمایه‌گذاری در سال‌های اخیر توسعه یافته است. در بخشی از این مطالعات از روش‌های یادگیری ماشین مثل رگرسیون و خوشه‌بندی استفاده شده است. در ادامه به مروری بر این مطالعات پرداخته شده است.

بیزر^{۵۵} و ساینی^{۵۶} (۲۰۱۵) در پژوهش خود به بررسی بازار سهام هند پرداختند. آن‌ها در این پژوهش از روش‌های K-Means، K-Medoids و Fast K-Means برای خوشه‌بندی ۵۰ نماد استفاده شده در شاخص نیفتی-۵۰ استفاده کردند. آن‌ها پس از خوشه‌بندی، خوشه‌های دارای بیش از ۳۰ نماد را حذف کردند و در

⁵⁴ Hierarchical Clustering

⁵⁵ Baser

⁵⁶ Saini

هر یک از خوشه‌های باقی‌مانده، از مدل کمترین واریانس مارکوویتز برای انتخاب سبد بهینه استفاده کردند. سپس از بین سبدهای سرمایه‌گذاری به دست آمده، بهترین سبد انتخاب و برای سرمایه‌گذاری پیشنهاد داده شد.

پژوهش دیگر در این حوزه مرتبط با چانگ^{۵۷} و همکاران (۲۰۱۶) است که ادعا می‌کند اولین پژوهشی است که از روش خوشه‌بندی انتشار وابستگی در مباحث اقتصادی استفاده کرده است. این پژوهش با استفاده از داده‌های قیمتی مربوط به ابتدای هر ماه از ۵۰۲ نماد موجود در شاخص S&P 500، اقدام به خوشه‌بندی با استفاده از بازده تاریخی و شاخص‌های تحلیل تکنیکال مانند RSI، KD، MACD و EMA کرده است. آن‌ها پس از این مرحله، نسبت به انتخاب تعداد ثابتی از نمادها بر اساس انحراف معیار به‌دست‌آمده اقدام کردند.

اروالو^{۵۸} و همکاران (۲۰۱۹) استفاده از خوشه‌بندی سلسله‌مراتبی را به عنوان رویکرد خود برگزیدند. آن‌ها با استفاده از هفت روش مختلف خوشه‌بندی سلسله‌مراتبی و مقایسه‌ی عملکرد آن‌ها با یکدیگر، بهترین روش را انتخاب و سپس نسبت به انتخاب سبد سرمایه‌گذاری در هر خوشه با استفاده از رویکرد بیشترین نسبت شارپ در روش مارکوویتز اقدام کردند.

میرلوحی و محمدی (۱۳۹۹) برای بررسی انتخاب سبد سرمایه‌گذاری در بورس اوراق بهادار تهران، از مدلی استفاده کردند که در آن نمادهای موجود در بازار بر اساس شاخص‌هایی مانند بازده، نسبت قیمت به درآمد (P/E)، شاخص ریسک سیستماتیک (β)، سود هر سهم (EPS) و نسبت ارزش بازار به ارزش دفتری (M/B) خوشه‌بندی و سپس بر اساس مدل مارکوویتز، در سبد سرمایه‌گذاری وزن‌دهی شدند.

⁵⁷ Chang
⁵⁸ Arévalo

بر خلاف دیگر پژوهش‌ها که یک سبد سرمایه‌گذاری از سهم‌های مختلف را بررسی می‌کند، تراز^{۵۹} و توکیو^{۶۰} (۲۰۲۱) در پژوهش خود یک سبد متشکل از صندوق‌های سرمایه‌گذاری ایجاد کردند که از روش میانگین-CVaR استفاده می‌کند. این روش یکی از روش‌های توسعه‌یافته‌ی مبتنی بر روش مارکوویتز است که به جای واریانس، از معیار ارزش در معرض خطر شرطی به عنوان سنجی ریسک استفاده می‌کند. آن‌ها برای خوشه‌بندی این صندوق‌های سرمایه‌گذاری، از روش خوشه‌بندی پویا بهره برده‌اند که یکی از روش‌های توسعه‌یافته‌ی مبتنی بر روش K-Means است.

ساس^{۶۱} و توس^{۶۲} (۲۰۲۱) از خوشه‌بندی برای بهبود روش هم‌وزن استفاده کردند. آن‌ها با بررسی بازار سهام آمریکا استدلال کردند که در روش هم‌وزن وقتی تعداد نمادها زیاد می‌شود، عملکرد سبد سرمایه‌گذاری افت می‌کند. مدل ارائه‌شده برای حل این مشکل، از روش‌های خوشه‌بندی سلسله‌مراتبی استفاده می‌کند تا با تقسیم نمادها به خوشه‌های متعدد، از هر خوشه یک نماد را که دارای بیشترین نسبت شارپ است به عنوان نماینده انتخاب کند و نمادهای نماینده را به با وزن یکسان در سبد سرمایه‌گذاری نهایی تقسیم کند. سلیمانی و وثیقی (۲۰۲۰) برای خوشه‌بندی داده‌های مرتبط با ۵۰۰ نماد بازار سهام نیویورک در آمریکا، از روش K-Means++ استفاده کردند. این روش یکی از روش‌های توسعه‌یافته مبتنی بر روش K-Means است که در آن نحوه‌ی یافتن مراکز خوشه‌ها متفاوت است و معمولاً عملکرد بهتری را ارائه می‌دهد. مدل استفاده‌شده در این مطالعه همچنین از شاخص‌های VaR و CVaR به عنوان شاخص ریسک استفاده می‌کند.

⁵⁹ Terraza

⁶⁰ Toque

⁶¹ Sass

⁶² Thös

پاندا^{۶۳} (۲۰۲۳) داده‌های مربوط به قیمت ماهانه‌ی نمادهای سه بازار سهام بزرگ آمریکا را در بازه‌ی زمانی بلندمدت ۱۹۶۵ تا ۲۰۲۲ بررسی و به روش سلسله‌مراتبی، آن‌ها را خوشه‌بندی کرده است. او سپس با استفاده از روش مارکوویتز اقدام به آماده‌سازی سبدهای سرمایه‌گذاری کرده که در آن‌ها فروش استقراضی نیز مجاز است. در نهایت، عملکرد سبدهای سرمایه‌گذاری با استفاده از سه معیار میانگین بازده، انحراف معیار و نسبت شارپ محاسبه شده است.

نوراحمدی و صادقی (۲۰۲۳) یک سبد سرمایه‌گذاری مبتنی بر خوشه‌بندی K-Means را پیشنهاد می‌کنند. آن‌ها برای خوشه‌بندی از لگاریتم بازده روزانه‌ی نمادها استفاده کردند و نماد دارای بیشترین نسبت شارپ را از هر یک از خوشه‌های حاصل‌شده انتخاب نمودند. در نهایت سبد سرمایه‌گذاری نهایی با استفاده از روش انتخاب سبد هم‌وزن از بین نمادهای باقی‌مانده به دست می‌آید.

رویکرد انتخاب یک نماینده برای هر خوشه در پژوهش داس^{۶۴} و همکاران (۲۰۲۳) نیز دیده می‌شود. آن‌ها از روش‌های مختلفی در هر یک از مراحل مدل خود استفاده کرده‌اند. در این پژوهش برای خوشه‌بندی، روش‌های K-Means، سلسله‌مراتبی، DBSCAN و انتشار وابستگی و برای انتخاب سبد سرمایه‌گذاری روش‌های هم‌وزن، متعادل‌سازی ریسک، میانگین-واریانس و روش داده‌محور استفاده شده است. آن‌ها برای انتخاب نماینده‌ی هر خوشه که به مرحله‌ی انتخاب سبد سرمایه‌گذاری می‌رود، از معیار بیشترین بازده استفاده کردند. آن‌ها در نهایت اقدام به مقایسه‌ی عملکرد روش‌های مختلف در ترکیب با یکدیگر نمودند تا بهترین روش‌های مرتبط با ترکیب خوشه‌بندی و انتخاب سبد سرمایه‌گذاری مشخص گردد.

⁶³ Panda

⁶⁴ Das

جدول ۱-۲: مقایسه‌ی مقالات

ردیف	مقاله	داده‌های مورد بررسی	روش خوشه‌بندی							روش انتخاب سبد		
			K-Means	K-Medoids	سلسله‌مراتبی	انتشار وابستگی	DBSCAN	سایر	ساده (EW)	مدل‌های مارکوویتز	مدل‌های توسعه‌یافته مبتنی بر مارکوویتز	سایر
۱	بیزر و ساینی (۲۰۱۵)	نمادهای شاخص نیفتی ۵۰ بازار سهام هند ۲۰۱۲-۰۱-۰۱ ۲۰۱۳-۱۲-۳۱	✓	✓	×	×	×	Fast K-Means	×	✓ (MVO)	×	×
۲	چانگ و همکاران (۲۰۱۶)	نمادهای شاخص S&P ۵۰۰ بازار سهام آمریکا ۲۰۰۱-۰۱-۰۱ ۲۰۱۵-۰۹-۳۱	×	×	×	✓	×	×	×	×	✓	×
۳	اروالو و همکاران (۲۰۱۹)	۲۰۰۰ نماد از بازار سهام آمریکا ۲۰۱۶-۰۱-۰۶ ۲۰۱۸-۰۱-۳۱	×	×	✓	×	×	×	×	✓ (SRO)	×	×
۴	میرلوحی و محمدی (۲۰۲۰)	نمادهای بورس اوراق بهادار تهران	✓	×	✓	×	×	×	×	✓	×	×
۵	ترازا و توکیو (۲۰۲۱)	۱۷ صندوق سرمایه‌گذاری بازار سهام لوکزامبورگ ۲۰۰۸-۰۱-۰۱ ۲۰۱۶-۱۲-۳۱	×	×	×	×	×	Dynamical Clustering	×	×	×	✓ (Mean-CVaR)

جدول ۲-۱: مقایسه‌ی مقالات (ادامه)

ردیف	مقاله	داده‌های مورد بررسی	روش خوشه‌بندی							روش انتخاب سبد		
			K-Means	K-Medoids	سلسله‌مراتبی	انتشار وابستگی	DBSCAN	سایر	ساده (EW)	مدل‌های مارکوویتز	مدل‌های توسعه‌یافته مبتنی بر مارکوویتز	سایر
۶	ساس و توس (۲۰۲۱)	نمادهای ۴۸ صنعت بازار سهام آمریکا ۱۹۶۹-۰۷-۰۱ تا ۲۰۱۴-۰۶-۳۰	×	×	✓	×	×	×	×	✓ (MVO)	×	×
۷	سلیمانی و وثیقی (۲۰۲۲)	۵۰۰ نماد بازار سهام نیویورک ۲۰۱۶-۰۱-۰۱ تا ۲۰۲۰-۰۳-۱۸	✓	×	×	×	×	×	×	×	✓ (Mean-VaR, Mean-CVaR)	×
۸	پاندا (۲۰۲۳)	بازار سهام آمریکا ۱۹۶۵-۰۱-۰۱ تا ۲۰۲۲-۱۲-۳۱	×	×	✓	×	×	×	×	✓	×	×
۹	نوراحمدی و صادقی (۲۰۲۳)	۳۳۴ نماد از بازار سهام ایران ۲۰۱۷-۰۱-۰۱ تا ۲۰۲۰-۰۷-۳۰	✓	×	×	×	×	×	×	✓	×	×
۱۰	داس و همکاران (۲۰۲۳)	بازار سهام آمریکا ۲۰۱۹-۰۱-۰۱ تا ۲۰۲۳-۱۲-۳۱	✓	×	✓	✓	✓	✓	×	✓ (MVO, SRO)	✓ (CPO)	IVW, DD
۱۱	پژوهش حاضر	۴۰ نماد دارای بیشترین ارزش در بازار رمزارزها ۲۰۲۱-۰۱-۰۱ تا ۲۰۲۴-۰۳-۰۱	✓	✓	×	✓	×	×	×	✓ (SRO)	✓ (Mean-CVaR)	HRP, Sparse

۲-۶- جمع‌بندی

انتخاب یک سبد سرمایه‌گذاری مناسب از دارایی‌های مورد نظر، همواره یکی از دغدغه‌های سرمایه‌گذاران بوده است. همان طور که مشاهده شد، روش‌های گوناگونی برای انتخاب و بهینه‌سازی سبدهای سرمایه‌گذاری توسعه یافته شده است تا سرمایه‌گذاران بتوانند بازدهی بیشتر و ریسک کمتری را تجربه کنند. پژوهشگران همچنین روش‌های غیربهمینه‌ای مثل روش هم‌وزن یا روش انتخاب پرتفوی بازار را در کنار روش‌های بهینه‌سازی مثل مدل مارکوویتز دنبال می‌کنند؛ چرا که هدف اصلی این است که سرمایه‌گذاران بتوانند در عمل از مزایای متنوع‌سازی دارایی‌ها بهره‌مند شوند. همچنین علاوه بر روش‌های ذکرشده، روش‌های جدیدی مانند مدل‌های مبتنی بر هوش مصنوعی به طور فزاینده در حال پیشرفت هستند و مقایسه‌ی این رویکردهای متفاوت با یکدیگر، از مطالعات لازم برای سرمایه‌گذاران است. در فصل آینده، پژوهش‌های انجام‌شده در حیطه‌ی انتخاب و بهینه‌سازی سبد دارایی‌ها در بازار رمزارزها مورد بررسی قرار می‌گیرد.

فصل سوم

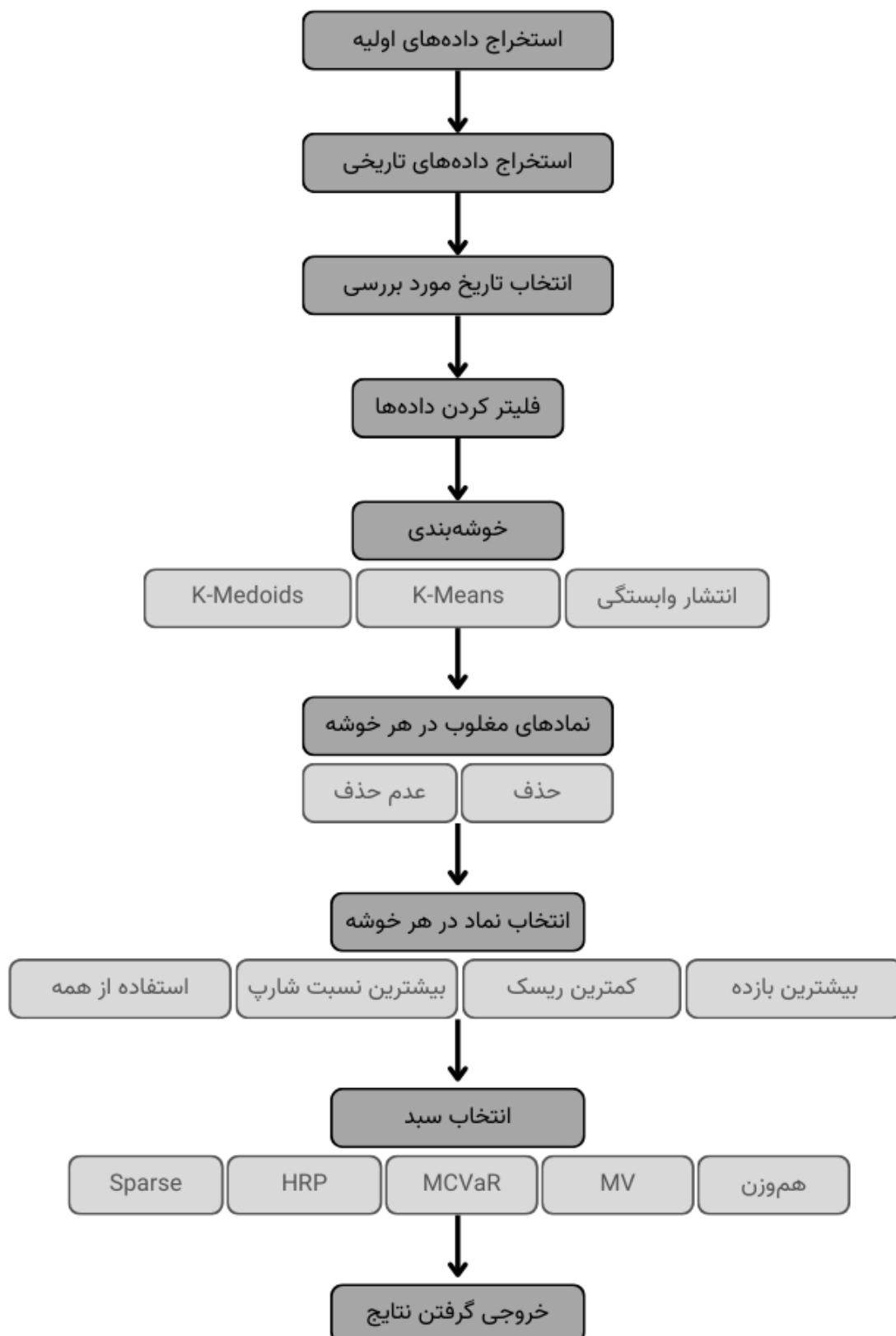
روش‌شناسی تحقیق

۳-۱- مقدمه

روش‌شناسی تحقیق که شامل مجموعه‌ای از قواعد و راه‌حل‌های معتبر است که پژوهش را به سرانجام می‌رساند، در این فصل مورد بررسی قرار می‌گیرد. مدل استفاده‌شده در این پژوهش از مراحل مختلفی تشکیل شده که به تفصیل در فصل حاضر شرح داده می‌شود.

۳-۲- روش تحقیق

پژوهش حاضر تلاش می‌کند بر اساس داده‌های نمونه‌ی جمع‌آوری‌شده، با روش‌های متفاوت در هر مرحله به اطلاعات نهایی وزن هر نماد در سبد سرمایه‌گذاری برسد و در نهایت بر اساس معیار نسبت شارپ، نتایج را با یکدیگر مقایسه و بهترین روش را انتخاب کند. این مراحل شامل استخراج و تمیزسازی داده‌ها، خوشه‌بندی داده‌ها، حذف نمادهای مغلوب در هر خوشه، انتخاب یک نماد در هر خوشه و در نهایت انتخاب سبد با استفاده از نمادهای باقی‌مانده است. شکل ۳-۱ این الگوریتم را به همراه روش‌های استفاده‌شده در هر مرحله نمایش می‌دهد. در ادامه‌ی این فصل، به تشریح هر یک از مراحل ذکرشده پرداخته می‌شود.



شکل ۳-۱: چارچوب روش تحقیق

۳-۳- استخراج داده‌های اولیه و تاریخی

همان طور که پیش از این ذکر شد، این پژوهش بازار رمزارزها را بررسی می‌کند. داده‌های مورد استفاده در این پژوهش از وبسایت کوین‌گکو^{۸۱} تهیه شده است. این وبسایت، یک رابط برنامه‌نویسی کاربردی (API) در اختیار مخاطبان قرار می‌دهد که اطلاعات اولیه و نیز اطلاعات تاریخی رمزارزها را شامل می‌شود.

در اولین مرحله از مدل حاضر، اطلاعات اولیه‌ی نمادها از طریق این API و توسط یک تابع نوشته‌شده با استفاده از زبان برنامه‌نویسی پایتون^{۸۲} بارگیری شده و برای استفاده آماده‌سازی می‌شود. تابع ایجادشده قابلیت محدودسازی نتایج را نیز دارد؛ به این شکل که می‌توان تعداد نمادها، بیشینه تاریخ اضافه شدن نماد، کمینه تاریخ بروز شدن نماد، و نیز کمینه بازار جفت رمزارزها را در نتایج نهایی محدود کرد. خروجی این مرحله یک لیست کامل از نمادهای فعال با فیلترهای اعمال شده خواهد بود.

مرحله‌ی بعد شامل بارگیری داده‌های تاریخی نمادها است. تابعی که این وظیفه را بر عهده دارد، از خروجی مرحله‌ی قبل استفاده می‌کند و برای هر یک نمادهای ذکرشده، داده‌های تاریخی روزانه را از ابتدای لیست شدن نماد تا آخرین روز بروز شدن آن بارگیری و ذخیره می‌نماید. این داده‌های تاریخی شامل قیمت باز شدن، قیمت بسته شدن، بیشترین قیمت، کمترین قیمت، حجم معاملات و ارزش بازار هر نماد در هر روز است. در ادامه دو مقدار قیمت میانگین و بازده روزانه نیز با استفاده از مقادیر قیمت باز شدن و قیمت بسته شدن در هر روز محاسبه و خروجی گرفته می‌شود.

⁸¹ coingecko.com

⁸² Python

۳-۴- فیلتر کردن داده‌ها بر اساس تاریخ

پس از استخراج داده‌های تاریخی، تاریخ بررسی داده‌ها انتخاب می‌شود. انتخاب تاریخ می‌تواند خود در چند مرحله انجام شود؛ به این شکل که برای صحت بیشتر مدل، بازه‌های زمانی متفاوتی مورد بررسی قرار گیرد و در نهایت از میانگین عملکرد مدل‌های مختلف استفاده گردد. پس از انتخاب تاریخ، بازه‌های زمانی مدل توسط یک تابع محدود می‌شود که ورودی آن، تاریخ جداسازی داده‌های گذشته، تعداد روزهای داده‌های گذشته، و تعداد روزهای داده‌های آینده است و بر این اساس، مدل به وسیله‌ی داده‌های گذشته اجرا شده و به وسیله‌ی داده‌های آینده، عملکردسنجی می‌شود. به این شکل می‌توان از عملکرد اعلام‌شده اطمینان حاصل کرد.

تابع ذکرشده، علاوه بر جداسازی داده‌های گذشته و آینده، نمادهایی را که شرایط مناسب برای بررسی ندارند حذف می‌کند. یکی از این شرایط این موضوع است که نماد در تمام روزهای معاملاتی، داده‌ی قیمتی داشته باشد. همچنین نمادهایی که در اولین و آخرین روز داده‌ی قیمتی نداشته باشند حذف می‌گردند. به این صورت، نمادهای جدیدتر که در تاریخ مورد بررسی هنوز شروع به فعالیت نکرده بودند از مدل کنار گذاشته می‌شوند. در نهایت خروجی این تابع، داده‌های تاریخی گذشته و داده‌های تاریخی آینده با توجه تاریخ داده‌شده و تعداد روزهای خواسته‌شده از مدل است.

۳-۵- خوشه‌بندی

پس از فیلتر کرده داده‌ها بر اساس شرایط مورد نیاز، از داده‌های گذشته که در مرحله‌ی قبل آماده شده است برای خوشه‌بندی نمادها استفاده می‌گردد. تقسیم داده‌ها به خوشه‌های مختلف به شکل نظارت‌نشده روش‌های متفاوتی دارد. مدل حاضر، به بررسی سه مورد از روش‌های خوشه‌بندی می‌پردازد که در ادامه توضیح داده خواهد شد.

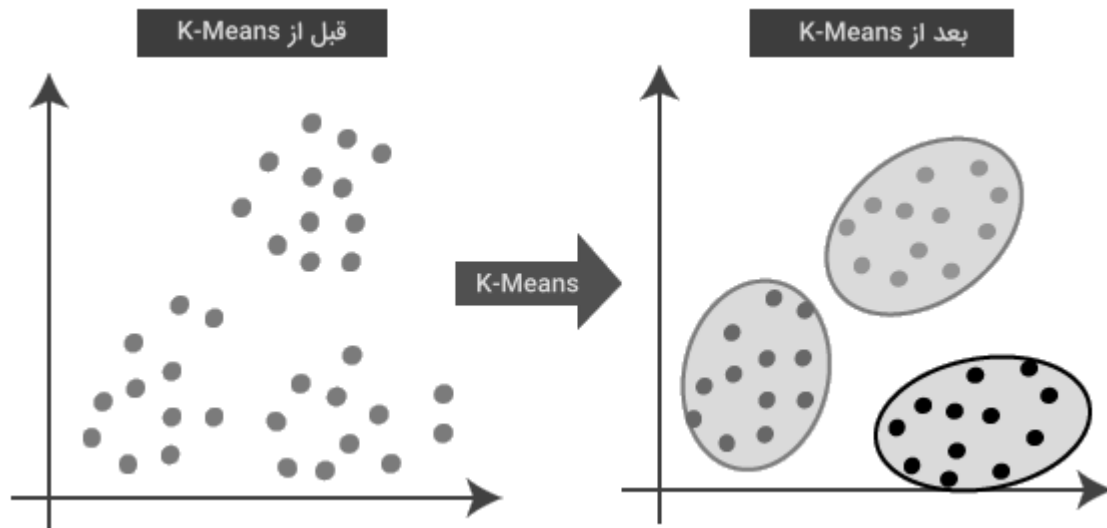
۳-۵-۱- خوشه‌بندی به روش انتشار وابستگی

انتشار وابستگی یکی از روش‌های خوشه‌بندی است که نسبت به دیگر روش‌ها متفاوت عمل می‌کند. این روش یک الگوریتم خوشه‌بندی بدون نیاز به تعیین تعداد خوشه است و در آن، همه‌ی نمونه‌ها با یکدیگر در ارتباط هستند و پس از شناسایی شباهت‌هایی که با هم دارند، فرایند خوشه‌بندی انجام می‌شود. همزمان با برقراری ارتباط، مجموعه‌ای از داده‌ها با عنوان نمونه‌های شاخص^{۸۳} پیدا شده که در واقع همان خوشه‌ها را تشکیل می‌دهند. یک نمونه‌ی شاخص زمانی پدیدار می‌شود که ارتباط میان نقاط داده شکل گرفته و بر سر بهترین نمونه‌ای که می‌تواند به عنوان خوشه شناخته شود، به توافق رسیده باشند.

۳-۵-۲- خوشه‌بندی به روش K-Means

یکی از شناخته‌ترین روش‌های خوشه‌بندی، روش K-Means است که با استفاده از یک فرایند تکرار کار می‌کند. این فرایند بهترین مقدار را برای نقاط مرکز تعیین میکند و هر نقطه‌ی داده را به نزدیک‌ترین مرکز خود اختصاص می‌دهد. به این شکل که در ابتدا بر اساس مقدار k ، تعدادی گروه با نقاط مرکزی تصادفی انتخاب می‌شوند و با محاسبه‌ی فاصله‌ی میان هر گروه و مراکز آن، داده‌هایی که به مرکز آن نزدیک‌تر هستند قرار می‌گیرند. سپس این فرآیند تا جایی تکرار می‌شود که موقعیت مرکز خوشه‌ها تغییر چندانی زیادی نداشته باشند و یا یک حد آستانه‌ی مشخصی رعایت شده باشد. نتیجه‌ی نهایی، چیزی مشابه شکل ۲-۳ خواهد بود.

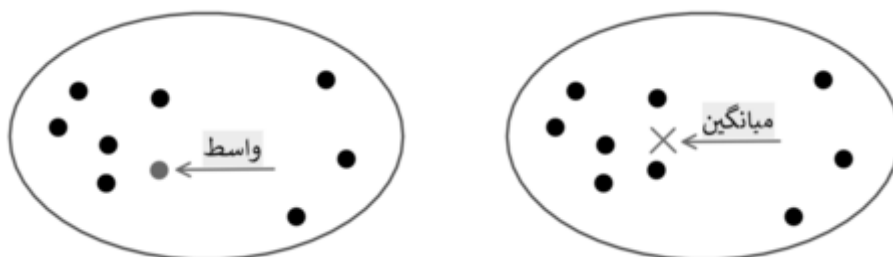
⁸³ Exemplars



شکل ۳-۲: مثالی از الگوریتم K-Means

۳-۵-۳ خوشه‌بندی به روش K-Medoids

این روش خوشه‌بندی شباهت زیادی با روش K-Means دارد و تنها تفاوت آن، استفاده از نقاط واسط به جای استفاده از نقاط میانگین است. همان طور که در شکل ۳-۳ مشاهده می‌شود، نقطه‌ی واسط عضوی از یک خوشه است که مجموع تفاوت‌هایش از دیگر اعضای آن مجموعه کمینه است. به این معنا که نقطه‌ی واسط حتماً عضوی از نقاط داده است در حالی که نقطه‌ی میانگین می‌تواند عضو از این مجموعه نباشد.



شکل ۳-۳: مثالی از الگوریتم K-Medoids

۳-۶- انتخاب نماد در هر خوشه

یکی از روش‌هایی بیشتر از روش‌های دیگر برای بهره‌گیری از خوشه‌بندی در انتخاب سبد استفاده می‌شود، این است که پس از خوشه‌بندی نمادها، یک یا چند نماد به عنوان نماینده‌ی هر خوشه در نظر گرفته شود و سایر نمادها از فرآیند انتخاب سبد حذف گردد. این هدف در دو مرحله در پژوهش حاضر پیاده‌سازی شده است که در ادامه ذکر می‌گردد.

در ابتدا مدل به دو صورت حذف یا عدم حذف نمادهای مغلوب در هر خوشه اقدام می‌نماید. حذف نمادهای مغلوب به این معناست که در داخل هر خوشه، تمامی نمادها از نظر بازده (با برآورد میانگین) و ریسک (با برآورد واریانس) به صورت دو به دو با یکدیگر مقایسه می‌شوند و نمادهایی که در هر مرحله از این مقایسه‌ها نسبت به نماد دیگری مغلوب شود، حذف می‌گردد. مغلوب شدن نماد این طور تعریف می‌شود که هم بازده پایین‌تری نسبت به نماد دیگر داشته باشد و هم ریسک آن بیشتر باشد.

با توجه به خروجی این مرحله، مدل وارد مرحله‌ی بعدی می‌شود. در این مرحله بر اساس روش مورد نظر یک نماد از بین همه‌ی نمادهای باقی‌مانده در هر خوشه انتخاب می‌شود. این نماد ممکن است بر اساس بیشترین بازده، کمترین ریسک، و یا بیشترین نسبت شارپ انتخاب شود. همچنین یک حالت دیگر هم وجود دارد که همه‌ی نمادهای باقی‌مانده از مرحله‌ی قبل به مرحله بعد یعنی مرحله‌ی انتخاب سبد منتقل شود.

نکته‌ی قابل توجه این است که اگر در هر دو مرحله‌ی این گام همه‌ی نمادها به مرحله‌ی بعد راه پیدا کنند، یعنی اگر هم نقاط مغلوب حذف نشوند و هم هیچ نمادی به عنوان نماینده انتخاب نشود، مانند این است که از خوشه‌بندی هیچ استفاده‌ای نشده است. به این ترتیب می‌توان در نهایت اثر خوشه‌بندی یا عدم خوشه‌بندی را بر انتخاب سبد سنجید.

۳-۷- انتخاب سبد

پس از فیلتر کردن نمادها در مراحل قبل، نوبت به انتخاب سبد سرمایه‌گذاری از بین این نمادهای باقی‌مانده می‌رسد. این کار به وسیله‌ی یک تابع انجام می‌گیرد که ورودی آن داده‌های تاریخی گذشته و مدل انتخاب سبد، و خروجی آن وزن نهایی هر نماد در سبد سرمایه‌گذاری است. مدل انتخاب سبد نیز همچون مراحل قبل متعدد و قابل انتخاب از بین یکی از روش‌هایی است که در ادامه توضیح داده خواهد شد.

۳-۷-۱- انتخاب سبد به روش هم‌وزن

روش هم‌وزن یکی از ساده‌ترین روش‌ها برای انتخاب سبد سرمایه‌گذاری است. در این روش، برای محاسبه‌ی وزن هر نماد در سبد سرمایه‌گذاری تنها کافی است که وزن یکسانی به تمام نمادها داده شود. بنابراین وزن‌دهی سبدي که شامل N نماد باشد، به این شکل انجام می‌شود که تمام نمادها وزن یکسان و ثابت $1/N$ را دریافت می‌کنند.

۳-۷-۲- انتخاب سبد به روش میانگین-واریانس

این روش که به آن روش مارکوویتز هم گفته می‌شود، بر اساس محاسبه و رسم مرز کارا در نمودار میانگین-واریانس عمل می‌کند. میانگین به عنوان برآوردی از بازده انتظاری و واریانس به عنوان برآوردی از ریسک انتظاری عمل می‌کند. همان‌طور که در فصل دوم توضیح داده شد، مرز کارا به این شکل به دست می‌آید که در نمودار میانگین-واریانس، تمام سبدهای سرمایه‌گذاری که حداقل یک سبد سرمایه‌گذاری دارای میانگین بیشتر و واریانس کمتر از آن‌ها وجود داشته باشد، حذف می‌گردد.

به بیان دیگر، با حذف سبدهای سرمایه‌گذاری مغلوب، مرزی از سبدهای سرمایه‌گذاری غالب شکل می‌گیرد که هیچ کدام بر یکدیگر برتری کامل ندارند و سرمایه‌گذار بسته به میزان ریسک‌پذیری خود، می‌تواند از بین آن‌ها سبد سرمایه‌گذاری مورد نظر را انتخاب کند.

برای انتخاب سبد سرمایه‌گذاری از بین نقاط موجود در مرز کارا، روش‌های مختلفی وجود دارد. یکی از این روش‌ها استفاده از نسبت شارپ است که از طریق رابطه‌ی ۲-۴ می‌توان آن را محاسبه و سبد سرمایه‌گذاری دارای بیشترین مقدار نسبت شارپ را به عنوان نقطه‌ی بهینه انتخاب کرد.

۳-۷-۳- انتخاب سبد به روش میانگین-CVaR^{۸۴}

این روش مانند روش مارکوویتز عمل می‌کند، با این تفاوت که به واریانس به عنوان معیاری برای سنجش ریسک سبد سرمایه‌گذاری، از شاخص ارزش در معرض ریسک شرطی استفاده می‌کند. این شاخص مقدار انتظاری زیان‌هایی را که فراتر از یک سطح معین رخ می‌دهند، اندازه‌گیری می‌کند.

به عبارت دیگر، CVaR به ارزیابی زیان‌های متوسط در بدترین درصد رخداد‌های ممکن می‌پردازد. این روش به ویژه در مواقعی که توزیع بازده‌ها نرمال نباشد کاربرد دارد. مقدار سطح اطمینان CVaR توسط استفاده‌کننده تعیین می‌شود. این مقدار در مدل فوق ۹۵ درصد در نظر گرفته شده که البته می‌توان آن را بر حسب شرایط تغییر داد.

⁸⁴ Conditional Value at Risk

۳-۷-۴- انتخاب سبد به روش متعادل سازی ریسک سلسله مراتبی یا HRP^{۸۵}

روش HRP یکی از روش های تعادل ریسک است که برای وزن دهی نمادها در سبد سرمایه گذاری، از تکنیک های ریاضیات مدرن، نظریه ی نمودار و یادگیری ماشین استفاده می کند و با توجه متعادل سازی ریسک نمادها در سبد سرمایه گذاری، منجر به عملکرد مناسب با ریسک بسیار پایین می شود. در الگوریتم HRP از ساختار طبقه بندی درختی استفاده می شود که به این صورت عمل می کند که نمادهای با شباهت بیشتر را در شاخه های پایین تر درخت قرار می دهد. این امر موجب می شود که بر خلاف مدل های دیگر مثل مدل مارکوویتز، همه ی نمادها جایگزین یکدیگر فرض نشوند و سبد سرمایه گذاری ثبات بیشتری داشته باشد.

۳-۷-۵- انتخاب سبد به روش تنک یا Sparse

در بهینه سازی، اصطلاح Sparse به حالتی اشاره دارد که بسیاری از عناصر یک بردار یا ماتریس صفر باشند. در زمینه ی بهینه سازی سبد سرمایه گذاری، این مفهوم به دنبال ایجاد یک سبد سرمایه گذاری است که فقط شامل تعداد محدودی از دارایی ها باشد. این کار باعث می شود که پرتفوی ساده تر و مدیریت آن آسان تر شود و هزینه های معامله و پیگیری نیز کاهش یابد.

در مدل حاضر انتخاب سبد به روش Sparse، به وسیله ی تابعی انجام می شود که تعداد معینی از نمادها که دارای شرایط بهتری هستند را نگه می دارد و وزن باقی نمادها را صفر می کند. در این روش که به آن انتخاب

⁸⁵ Hierarchical Risk Parity

مجموعه^{۸۶} گفته می‌شود، به جای بهینه‌سازی سبد سرمایه‌گذاری با تمام دارایی‌ها، زیرمجموعه‌ای از دارایی‌ها انتخاب می‌شود که بهترین عملکرد را دارند.

۳-۸- خروجی گرفتن نتایج

پس از مشخص شدن وزن هر نماد در سبد سرمایه‌گذاری، لازم است که عملکرد سبد نهایی بررسی شود. با توجه به تقسیم داده‌ها به داده‌های گذشته و آینده در مدل و استفاده از داده‌های گذشته تا به اینجای کار، برای بررسی عملکرد سبد سرمایه‌گذاری می‌بایست از داده‌های آینده استفاده شود تا ارزیابی سبد بدون سوگیری نتایج صورت پذیرد.

تابعی که وظیفه‌ی ارزیابی عملکرد سبد سرمایه‌گذاری را بر عهده دارد، پرتفولیوی نهایی مدل و داده‌های قیمتی مربوط به بازه‌ی آینده را دریافت می‌کند و بررسی می‌کند که اگر یک سرمایه‌گذار با توجه به وزن‌های پیشنهادی مدل اقدام به سرمایه‌گذاری می‌کرد، چه مقدار سود یا زیان و با چه ریسکی به دست می‌آورد.

می‌دانیم که برای بررسی عملکرد سبد سرمایه‌گذاری لازم است به هر دو پارامتر بازده و ریسک به طور همزمان توجه شود. یکی از روش‌های پذیرفته‌شده برای ارزیابی یک سرمایه‌گذاری که به هر دو پارامتر فوق توجه می‌کند، نسبت شارپ است که رابطه‌ی ۳-۱ بیانگر نحوه‌ی محاسبه‌ی آن است. در نهایت هرچه یک سبد سرمایه‌گذاری نسبت شارپ بیشتری داشته باشد، عملکرد بهتری داشته است.

در نهایت با مشخص شدن عملکرد هر یک از سبدهای سرمایه‌گذاری با استفاده از هر کدام از روش‌های ذکرشده در هر مرحله، خروجی نهایی برای مقایسه‌ی هر روش گرفته می‌شود.

⁸⁶ Subset Selection

۳-۹- جمع‌بندی

در این فصل به بررسی روش تحقیق و ارائه‌ی مدل پیشنهادی برای بهینه‌سازی سبد سرمایه‌گذاری پرداخته شد. این مدل شامل مراحل مختلف و همچنین روش‌های مختلف در هر مرحله است تا با توجه به عملکرد نهایی هر یک از روش‌های آزمون‌شده روی داده‌های گذشته، سرمایه‌گذار بتواند بهترین روش را انتخاب کند و برای استفاده در سرمایه‌گذاری‌های آتی از آن استفاده نماید. در فصل بعد به پیاده‌سازی این مدل و مرور نتایج حاصل‌شده از آن پرداخته می‌شود.

فصل چهارم

پیاده‌سازی و نتایج پژوهش

۴-۱- مقدمه

این فصل به پیاده‌سازی مدل شرح داده شده در فصل سوم و تحلیل نتایج مربوط به آن می‌پردازد. همان طور که پیش از این ذکر شد، تمام مراحل این پیاده‌سازی با استفاده از زبان برنامه‌نویسی پایتون و کتابخانه‌های مرتبط انجام شده است. همچنین برای تحلیل نتایج، خروجی نهایی مدل در قالب فایل اکسل^{۸۷} به عنوان ورودی به نرم‌افزار تبلو^{۸۸} ارائه و نمودارهای مورد نیاز، در این نرم‌افزار رسم شده است. این نمودارها در نهایت مقایسه‌ای از بین روش‌های موجود در مدل ارائه می‌کند تا سرمایه‌گذار بتواند بهترین این روش‌ها را بر حسب نیاز خود انتخاب و استفاده کند.

۴-۲- داده‌های ورودی

داده‌های ورودی این پژوهش از وبسایت کوین‌گکو به دست آمده است. فراخوانی API این وبسایت در دو مرحله انجام می‌شود. مرحله‌ی نخست، دریافت داده‌های کلی مربوط به تمام رمزارزهای بازار است که اطلاعات کلی نمادها را ارائه می‌دهد. در این بخش از مدل، اطلاعات کلی مرتبط با ۴۰ نماد نخست از نظر ارزش بازار دریافت می‌گردد. همچنین اطلاعات مربوط به نمادهایی که از سال ۲۰۱۹ به بعد اضافه و نیز

⁸⁷ Excel

⁸⁸ Tableau

نمادهایی که از سال ۲۰۲۴ به بعد بروزرسانی نشده اند، از خروجی تابع حذف می‌گردد. شرط دیگری که در این بخش اعمال شده، مربوط به حداقل جفت‌ارزهای معاملاتی در صرافی‌ها است که این مقدار برابر با ۶ عدد قرار داده شده است. شایان ذکر است که در تابع ذکرشده، می‌توان هر یک از این شرط‌ها را بسته به خواسته‌ی سرمایه‌گذاران تغییر داد.

name	symbol	slug	marketPairCount	isActive	lastUpdated	dateAdded
Bitcoin	BTC	bitcoin	11077	1	2024-05-30T10:35:00.000Z	2010-07-13T00:00:00.000Z
Ethereum	ETH	ethereum	9012	1	2024-05-30T10:34:00.000Z	2015-08-07T00:00:00.000Z
Tether USDt	USDT	tether	87194	1	2024-05-30T10:34:00.000Z	2015-02-25T00:00:00.000Z
BNB	BNB	bnb	2172	1	2024-05-30T10:34:00.000Z	2017-07-25T00:00:00.000Z
USDC	USDC	usd-coin	19490	1	2024-05-30T10:34:00.000Z	2018-10-08T00:00:00.000Z
XRP	XRP	xrp	1331	1	2024-05-30T10:35:00.000Z	2013-08-04T00:00:00.000Z
Dogecoin	DOGE	dogecoin	974	1	2024-05-30T10:35:00.000Z	2013-12-15T00:00:00.000Z
Cardano	ADA	cardano	1186	1	2024-05-30T10:34:00.000Z	2017-10-01T00:00:00.000Z
Chainlink	LINK	chainlink	1784	1	2024-05-30T10:34:00.000Z	2017-09-20T00:00:00.000Z

شکل ۴-۱: نمونه‌ای از داده‌های ورودی در مرحله‌ی نخست

مرحله‌ی دوم فراخوانی و دریافت داده‌ها، مربوط به دریافت داده‌های تاریخی قیمت هر یک نمادها است. در این مرحله اطلاعات قیمتی تمامی نمادهایی که از مرحله‌ی قبل عبور کرده اند دریافت و ذخیره می‌شود.

symbol	date	open	high	low	close	avg	return	volume	marketcap
BTC	2024-04-26	64,485	64,790	63,322	63,755	64,120	-1.1%	24,139,372,950	1,255,299,007,880
BTC	2024-04-25	64,275	65,275	62,784	64,482	64,378	0.3%	32,155,786,816	1,269,733,248,757
BTC	2024-04-24	66,409	67,075	63,590	64,277	65,343	-3.2%	30,276,655,120	1,265,561,280,453
BTC	2024-04-23	66,840	67,199	65,865	66,407	66,624	-0.6%	24,310,975,583	1,307,512,895,862
BTC	2024-04-22	64,936	67,234	64,548	66,838	65,887	2.9%	28,282,686,673	1,315,994,975,669
BTC	2024-04-21	64,993	65,723	64,278	64,927	64,960	-0.1%	20,506,644,853	1,278,296,008,665
BTC	2024-04-20	63,851	65,442	63,172	64,994	64,423	1.8%	23,097,485,495	1,279,569,409,422
BTC	2024-04-19	63,511	65,482	59,651	63,844	63,677	0.5%	49,920,425,401	1,257,063,743,135
BTC	2024-04-18	61,275	64,126	60,833	63,513	62,394	3.7%	36,006,307,335	1,250,301,972,764

شکل ۴-۲: نمونه‌ای از داده‌های ورودی در مرحله‌ی دوم

پس از ذخیره‌سازی داده‌های قیمتی، سرمایه‌گذار می‌تواند انتخاب کند که مدل بر اساس داده‌های چند روز گذشته تمرین داده شود (داده‌های گذشته) و ارزیابی عملکرد نهایی مدل بر اساس داده‌های چند روز آتی

باشد (داده‌های آینده). بنابراین افق سرمایه‌گذاری در این بخش از مدل مشخص می‌شود. پژوهش حاضر، داده‌های گذشته را بر اساس ۳۶۰ روز پیش از هر تاریخ، و داده‌های آینده را بر اساس ۶۰ روز پس از هر تاریخ تقسیم‌بندی می‌کند.

همچنین برای اطمینان از نتایج و در نظر گرفتن همه‌ی شرایط بازار، تنها به آزمون مدل در یک تاریخ خاص اکتفا نشده است. پژوهش حاضر به بررسی ۲۰ تاریخ متفاوت، از ابتدای سال ۲۰۲۱ تا سومین ماه سال ۲۰۲۴ به فاصله‌ی هر دو ماه یک بار اجرا و میانگین نتایج با یکدیگر مقایسه شده است.

۴-۳- پیاده‌سازی مدل

همان طور که در بخش روش‌شناسی تحقیق توضیح داده شد، پس از دریافت داده‌ها و انتخاب تاریخ اجرای مدل، نوبت به مرحله‌ی خوشه‌بندی نمادها بر اساس داده‌های تاریخی می‌رسد. این مرحله با استفاده از سه روش انتشار وابستگی، K-Means و K-Medoids انجام می‌شود.

تعداد خوشه‌ها در روش انتشار وابستگی به صورت خودکار و توسط مدل انتخاب می‌شود، اما در دو روش دیگر باید تعداد خوشه‌ها توسط مدلساز تعیین شود. برای دریافت بهترین نتیجه، یک عدد حداکثری برای تعداد خوشه‌ها تعیین می‌شود و مدل برای دو الی عدد تعیین شده اجرا می‌شود و بهترین نتیجه از بین نتایج هر یک از روش‌ها انتخاب می‌گردد. این عدد در پژوهش حاضر ۱۰ خوشه تعیین شده است. بهترین نتیجه‌ی خوشه‌بندی بر اساس شاخصی به نام شاخص نیم‌رخ یا سایه‌نما^{۸۹} تعیین می‌شود که معیاری برای اندازه‌گیری پیوستگی درون خوشه‌ها و تفکیک‌پذیری آن‌هاست.

⁸⁹ Silhouette

symbol	cluster
DOGE	0
BTC	1
ETH	1
LTC	1
MATIC	1
TRX	1
XRP	1
USDC	2
USDT	2

شکل ۴-۳: نمونه‌ای از نتیجه‌ی خوشه‌بندی نمادها

پس از تقسیم شدن نمادها در خوشه‌ها در هر روش، نماد یا نمادهایی درون هر خوشه برای مراحل بعدی انتخاب می‌شود. انتخاب نمادها در دو مرحله انجام می‌شود. در مرحله‌ی نخست سرمایه‌گذار می‌تواند انتخاب کند که نمادهای در هر خوشه حذف شوند. در مرحله‌ی دوم نیز، سرمایه‌گذار می‌تواند انتخاب کند که یک نماد از بین نمادهای هر خوشه به عنوان نماینده‌ی آن خوشه به مرحله‌ی بعد برود. این انتخاب می‌تواند بر اساس معیارهایی چون بیشترین بازده، کمترین ریسک یا بیشترین نسبت شارپ باشد. در صورتی که سرمایه‌گذار در هر دو مرحله تصمیم بگیرد که هیچ نمادی حذف نشود، گویی از خوشه‌بندی نمادها هیچ استفاده‌ای نشده است. این حالت برای ارزیابی تأثیر انجام خوشه‌بندی بر نتیجه‌ی نهایی مؤثر است.

پس از فیلتر شدن نمادها در مراحل قبل، همه‌ی نمادهای باقی‌مانده به مرحله‌ی بهینه‌سازی سبد سرمایه‌گذاری منتقل می‌شوند. در این مرحله وزن هر یک از نمادها بر اساس روش‌هایی مانند روش هم‌وزن، روش مارکوویتز (با معیار واریانس و یا ارزش در معرض ریسک شرطی)، روش HRP و روش Sparse در یک سبد سرمایه‌گذاری تعیین می‌شود. در نهایت نیز با سنجیدن بازده، ریسک و نسبت شارپ سبد سرمایه‌گذاری به‌دست‌آمده، نتایج حاصل‌شده تحلیل می‌شود.

end_date	clustering_method	use_domination	symbol_selection_method	portfolio_selection_method	portfolio_return	portfolio_risk	sharpe_ratio
2021-01-01	affinity_propagation	TRUE	max_return	ew	0.007241624	0.03685048	0.1965137
2021-01-01	affinity_propagation	TRUE	max_return	mv	0.001134529	0.01895503	0.0598537
2021-01-01	affinity_propagation	TRUE	max_return	mcvar	0.001563487	0.03521925	0.04439296
2021-01-01	affinity_propagation	TRUE	max_return	hrp	0.000401909	0.00208363	0.192889
2021-01-01	affinity_propagation	TRUE	max_return	sparse	0.005848352	0.02263046	0.2584283
2021-01-01	affinity_propagation	TRUE	min_risk	ew	0.007081003	0.03578693	0.19786559
2021-01-01	affinity_propagation	TRUE	min_risk	mv	0.003583592	0.02145984	0.16699057
2021-01-01	affinity_propagation	TRUE	min_risk	mcvar	0.013077519	0.03515663	0.37197879
2021-01-01	affinity_propagation	TRUE	min_risk	hrp	0.000430177	0.00207043	0.20777125
2021-01-01	affinity_propagation	TRUE	min_risk	sparse	0.005781205	0.02245604	0.25744544
2021-01-01	affinity_propagation	TRUE	max_sharpe	ew	0.007241624	0.03685048	0.1965137
2021-01-01	affinity_propagation	TRUE	max_sharpe	mv	0.001134529	0.01895503	0.0598537
2021-01-01	affinity_propagation	TRUE	max_sharpe	mcvar	0.001563487	0.03521925	0.04439296
2021-01-01	affinity_propagation	TRUE	max_sharpe	hrp	0.000401909	0.00208363	0.192889
2021-01-01	affinity_propagation	TRUE	max_sharpe	sparse	0.005848352	0.02263046	0.2584283

شکل ۴-۴: نمونه‌ای از ارزیابی عملکرد سبدهای سرمایه‌گذاری

۴-۴- نتایج مدل

همان‌طور که در شکل ۴-۴ مشاهده می‌شود، نتایج نهایی مدل در قالب جدولی شامل تاریخ، روش استفاده‌شده در هر مرحله و معیارهای ارزیابی عملکرد حاصل‌شده ارائه می‌گردد. برای مقایسه‌ی روش‌ها، لازم است از معیارهای ارزیابی عملکرد یعنی بازده، ریسک و نسبت شارپ سبد سرمایه‌گذاری در تاریخ‌های مختلف اجرای مدل میانگین گرفته شود. نتایج نهایی مدل پس از میانگین‌گیری در جدول ۴-۱ قابل مشاهده است.

جدول ۴-۱: نتایج مدل

روش خوشه‌بندی	روش انتخاب نماد در هر خوشه	روش انتخاب سبد	نقاط مغلوب	بازده	ریسک	نسبت شارپ
K-Means	بیشترین شارپ	HRP	حذف	۰.۲۸٪	۰.۰۵۸۹	۱۱.۱۷٪
K-Means	بیشترین شارپ	HRP	عدم حذف	۰.۲۸٪	۰.۰۵۸۹	۱۱.۱۷٪
K-Means	بیشترین شارپ	Sparse	حذف	۰.۴۰٪	۰.۰۶۱۴	۱۰.۸۳٪
K-Means	بیشترین شارپ	Sparse	عدم حذف	۰.۴۰٪	۰.۰۶۱۴	۱۰.۸۳٪
K-Means	بیشترین شارپ	هم‌وزن	حذف	۰.۴۳٪	۰.۰۵۹۱	۱۰.۸۲٪
K-Means	بیشترین شارپ	هم‌وزن	عدم حذف	۰.۴۳٪	۰.۰۵۹۱	۱۰.۸۲٪
K-Means	بیشترین بازده	Sparse	حذف	۰.۳۸٪	۰.۰۵۹	۱۰.۵۹٪
K-Means	بیشترین بازده	Sparse	عدم حذف	۰.۳۸٪	۰.۰۵۹	۱۰.۵۹٪
K-Means	بیشترین بازده	HRP	حذف	۰.۲۷٪	۰.۰۵۶	۱۰.۴۹٪
K-Means	بیشترین بازده	HRP	عدم حذف	۰.۲۷٪	۰.۰۵۶	۱۰.۴۹٪
انتشار وابستگی	بیشترین شارپ	هم‌وزن	حذف	۰.۲۸٪	۰.۰۳۵۷	۱۰.۴۱٪
انتشار وابستگی	بیشترین شارپ	هم‌وزن	عدم حذف	۰.۲۸٪	۰.۰۳۵۷	۱۰.۴۱٪
K-Means	بیشترین بازده	هم‌وزن	حذف	٪۰.۴۰	۰.۰۵۶۶	٪۱۰.۳۹
K-Means	بیشترین بازده	هم‌وزن	عدم حذف	٪۰.۴۰	۰.۰۵۶۶	٪۱۰.۳۹
انتشار وابستگی	بیشترین شارپ	Sparse	حذف	٪۰.۲۱	۰.۰۲۶۳	٪۹.۸۸
انتشار وابستگی	بیشترین شارپ	Sparse	عدم حذف	٪۰.۲۱	۰.۰۲۶۳	٪۹.۸۸
انتشار وابستگی	کمترین ریسک	هم‌وزن	حذف	٪۰.۲۴	۰.۰۳۰۳	٪۹.۷۴
انتشار وابستگی	کمترین ریسک	هم‌وزن	عدم حذف	٪۰.۲۴	۰.۰۳۰۳	٪۹.۷۴
انتشار وابستگی	بیشترین بازده	هم‌وزن	حذف	٪۰.۲۳	۰.۰۳۱۹	٪۹.۷۳
انتشار وابستگی	بیشترین بازده	هم‌وزن	عدم حذف	٪۰.۲۳	۰.۰۳۱۹	٪۹.۷۳
K-Means	کمترین ریسک	Sparse	حذف	٪۰.۳۱	۰.۰۳۱۷	٪۹.۶۲
K-Means	کمترین ریسک	Sparse	عدم حذف	٪۰.۳۱	۰.۰۳۱۷	٪۹.۶۲
انتشار وابستگی	بیشترین بازده	Sparse	حذف	٪۰.۱۸	۰.۰۲۲۱	٪۹.۶۲
انتشار وابستگی	بیشترین بازده	Sparse	عدم حذف	٪۰.۱۸	۰.۰۲۲۱	٪۹.۶۲
K-Means	استفاده از همه	هم‌وزن	حذف	٪۰.۲۲	۰.۰۳۵۲	٪۹.۵۷
K-Means	کمترین ریسک	میانگین-CVaR	حذف	٪۰.۶۳	۰.۰۵۱۴	٪۹.۵۴
K-Means	کمترین ریسک	میانگین-CVaR	عدم حذف	٪۶۳.۰	۰.۰۵۱۴	٪۹.۵۴
K-Means	کمترین ریسک	هم‌وزن	حذف	٪۰.۳۶	۰.۰۳۷۳	٪۹.۴۹
K-Means	کمترین ریسک	هم‌وزن	عدم حذف	٪۰.۳۶	۰.۰۳۷۳	٪۹.۴۹
K-Means	بیشترین شارپ	میانگین-CVaR	حذف	٪۰.۵۹	۰.۰۸۲۱	٪۹.۳۵
K-Means	بیشترین شارپ	میانگین-CVaR	عدم حذف	٪۰.۵۹	۰.۰۸۲۱	٪۹.۳۵
انتشار وابستگی	استفاده از همه	هم‌وزن	حذف	٪۰.۲۳	۰.۰۳۳۵	٪۹.۱۱
K-Medoids	استفاده از همه	هم‌وزن	حذف	٪۰.۲۳	۰.۰۳۲۳	٪۸.۷۶
K-Means	بیشترین بازده	میانگین-CVaR	حذف	٪۰.۵۲	۰.۰۷۸۶	٪۸.۵۲

جدول ۴-۱: نتایج مدل (ادامه)

روش خوشه‌بندی	روش انتخاب نماد در هر خوشه	روش انتخاب سبد	نقاط مغلوب	بازده	ریسک	نسبت شارپ
K-Means	بیشترین بازده	میانگین-CVaR	عدم حذف	٪۰.۵۲	۰.۰۷۸۶	٪۸.۵۲
K-Means	استفاده از همه	Sparse	حذف	٪۰.۱۳	۰.۰۲۵۳	٪۸.۴۳
انتشار وابستگی	کمترین ریسک	Sparse	حذف	٪۰.۱۳	۰.۰۲۶۵	٪۸.۱۹
انتشار وابستگی	کمترین ریسک	Sparse	عدم حذف	٪۰.۱۳	۰.۰۲۶۵	٪۸.۱۹
K-Medoids	استفاده از همه	Sparse	حذف	٪۰.۱۶	۰.۰۲۳۴	٪۷.۸۸
K-Medoids	استفاده از همه	هم‌وزن	عدم حذف	٪۰.۲۳	۰.۰۳۶۸	٪۷.۸۷
K-Means	استفاده از همه	هم‌وزن	عدم حذف	٪۰.۲۳	۰.۰۳۶۸	٪۷.۸۷
انتشار وابستگی	استفاده از همه	هم‌وزن	عدم حذف	٪۰.۲۳	۰.۰۳۶۸	٪۷.۸۷
K-Medoids	بیشترین شارپ	HRP	حذف	٪۰.۴۶	۰.۰۸۱	٪۷.۳۲
K-Medoids	بیشترین شارپ	HRP	عدم حذف	٪۰.۴۶	۰.۰۸۱	٪۷.۳۲
K-Medoids	بیشترین شارپ	Sparse	حذف	٪۰.۴۵	۰.۰۷۵۷	٪۷.۰۴
K-Medoids	بیشترین شارپ	Sparse	عدم حذف	٪۰.۴۵	۰.۰۷۵۷	٪۷.۰۴
K-Medoids	بیشترین شارپ	هم‌وزن	حذف	٪۰.۴۷	۰.۰۷۲	٪۷.۰۲
K-Medoids	بیشترین شارپ	هم‌وزن	عدم حذف	٪۰.۴۷	۰.۰۷۲	٪۷.۰۲
K-Medoids	کمترین ریسک	میانگین-واریانس	حذف	٪۰.۳۱	۰.۰۵۹۹	٪۶.۵۸
K-Medoids	کمترین ریسک	میانگین-واریانس	عدم حذف	٪۰.۳۱	۰.۰۵۹۹	٪۶.۵۸
K-Medoids	بیشترین بازده	Sparse	حذف	٪۰.۴۱	۰.۰۷۲۲	٪۶.۴۵
K-Medoids	بیشترین بازده	Sparse	عدم حذف	٪۰.۴۱	۰.۰۷۲۲	٪۶.۴۵
K-Medoids	بیشترین بازده	هم‌وزن	حذف	٪۰.۴۲	۰.۰۶۸۸	٪۶.۴۴
K-Medoids	بیشترین بازده	هم‌وزن	عدم حذف	٪۰.۴۲	۰.۰۶۸۸	٪۶.۴۴
K-Medoids	بیشترین بازده	HRP	حذف	٪۰.۴۱	۰.۰۷۴	٪۶.۳۷
K-Medoids	بیشترین بازده	HRP	عدم حذف	٪۰.۴۱	۰.۰۷۴	٪۶.۳۷
K-Means	استفاده از همه	میانگین-CVaR	حذف	٪۰.۲۱	۰.۰۳۲۶	٪۶.۲۹
K-Means	کمترین ریسک	میانگین-واریانس	حذف	٪۰.۲۵	۰.۰۴۹	٪۵.۸۷
K-Means	کمترین ریسک	میانگین-واریانس	عدم حذف	٪۰.۲۵	۰.۰۴۹	٪۵.۸۷
انتشار وابستگی	کمترین ریسک	میانگین-واریانس	حذف	٪۰.۱۲	۰.۰۳۸	٪۵.۸۶
انتشار وابستگی	کمترین ریسک	میانگین-واریانس	عدم حذف	٪۰.۱۲	۰.۰۳۸	٪۵.۸۶
K-Medoids	استفاده از همه	میانگین-CVaR	حذف	٪۰.۳۰	۰.۰۳۵۹	٪۵.۶۴
K-Medoids	استفاده از همه	میانگین-واریانس	حذف	٪۰.۱۷	۰.۰۳۸۵	٪۵.۶۲
K-Medoids	کمترین ریسک	Sparse	حذف	٪۰.۱۴	۰.۰۲۲۶	٪۵.۲۵
K-Medoids	کمترین ریسک	Sparse	عدم حذف	٪۰.۱۴	۰.۰۲۲۶	٪۵.۲۵
K-Medoids	کمترین ریسک	هم‌وزن	حذف	٪۰.۱۵	۰.۰۲۴۴	٪۵.۲۴
K-Medoids	کمترین ریسک	هم‌وزن	عدم حذف	٪۰.۱۵	۰.۰۲۴۴	٪۵.۲۴
K-Means	کمترین ریسک	HRP	حذف	٪۰.۰۰	۰.۰۰۰۹	٪۵.۲۱

جدول ۴-۱: نتایج مدل (ادامه)

روش خوشه‌بندی	روش انتخاب نماد در هر خوشه	روش انتخاب سبد	نقاط مغلوب	بازده	ریسک	نسبت شارپ
K-Means	کمترین ریسک	HRP	عدم حذف	٪۰.۰۰	۰.۰۰۰۹	٪۵.۲۱
انتشار وابستگی	بیشترین شارپ	میانگین-واریانس	حذف	٪۰.۱۳	۰.۰۴۱۴	٪۵.۱۸
انتشار وابستگی	بیشترین شارپ	میانگین-واریانس	عدم حذف	٪۰.۱۳	۰.۰۴۱۴	٪۵.۱۸
K-Medoids	استفاده از همه	میانگین-واریانس	عدم حذف	٪۰.۱۵	۰.۰۳۶۴	٪۵.۱۲
K-Means	استفاده از همه	میانگین-واریانس	عدم حذف	٪۰.۱۵	۰.۰۳۶۴	٪۵.۱۲
انتشار وابستگی	استفاده از همه	میانگین-واریانس	عدم حذف	٪۰.۱۵	۰.۰۳۶۴	٪۵.۱۲
انتشار وابستگی	استفاده از همه	میانگین-واریانس	حذف	٪۰.۱۳	۰.۰۳۷۴	٪۴.۵۲
K-Means	استفاده از همه	میانگین-واریانس	حذف	٪۰.۱۴	۰.۰۳۵۷	٪۴.۴۹
K-Means	استفاده از همه	HRP	حذف	٪۰.۰۱	۰.۰۰۰۸	٪۴.۴۹
K-Medoids	استفاده از همه	Sparse	عدم حذف	٪۰.۰۹	۰.۰۲۸۳	٪۴.۱۰
K-Means	استفاده از همه	Sparse	عدم حذف	٪۰.۰۹	۰.۰۲۸۳	٪۴.۱۰
انتشار وابستگی	استفاده از همه	Sparse	عدم حذف	٪۰.۰۹	۰.۰۲۸۳	٪۴.۱۰
K-Medoids	کمترین ریسک	میانگین-CVaR	حذف	٪۰.۱۵	۰.۰۲۹	٪۴.۰۶
K-Medoids	کمترین ریسک	میانگین-CVaR	عدم حذف	٪۰.۱۵	۰.۰۲۹	٪۴.۰۶
انتشار وابستگی	استفاده از همه	Sparse	حذف	٪۰.۰۰	۰.۰۲۶۲	٪۴.۰۴
انتشار وابستگی	بیشترین بازده	میانگین-واریانس	حذف	٪۰.۱۱	۰.۰۳۹۹	٪۳.۸۵
انتشار وابستگی	بیشترین بازده	میانگین-واریانس	عدم حذف	٪۰.۱۱	۰.۰۳۹۹	٪۳.۸۵
انتشار وابستگی	بیشترین شارپ	میانگین-CVaR	حذف	٪۰.۱۳	۰.۰۳۵۲	٪۳.۶۸
انتشار وابستگی	بیشترین شارپ	میانگین-CVaR	عدم حذف	٪۰.۱۳	۰.۰۳۵۲	٪۳.۶۸
K-Medoids	بیشترین شارپ	میانگین-CVaR	حذف	٪۰.۳۳	۰.۰۹۱۵	٪۳.۶۸
K-Medoids	بیشترین شارپ	میانگین-CVaR	عدم حذف	٪۰.۳۳	۰.۰۹۱۵	٪۳.۶۸
K-Medoids	استفاده از همه	HRP	حذف	٪۰.۰۱	۰.۰۰۰۹	٪۳.۶۲
K-Medoids	بیشترین بازده	میانگین-CVaR	حذف	٪۰.۳۲	۰.۰۸۷۶	٪۳.۵۶
K-Medoids	بیشترین بازده	میانگین-CVaR	عدم حذف	٪۰.۳۲	۰.۰۸۷۶	٪۳.۵۶
K-Medoids	بیشترین شارپ	میانگین-واریانس	حذف	٪۰.۲۴	۰.۰۸۴۵	٪۳.۵۳
K-Medoids	بیشترین شارپ	میانگین-واریانس	عدم حذف	٪۰.۲۴	۰.۰۸۴۵	٪۳.۵۳
K-Medoids	بیشترین بازده	میانگین-واریانس	حذف	٪۰.۲۳	۰.۰۸۱۲	٪۳.۴۵
K-Medoids	بیشترین بازده	میانگین-واریانس	عدم حذف	٪۰.۲۳	۰.۰۸۱۲	٪۳.۴۵
انتشار وابستگی	استفاده از همه	HRP	حذف	٪۰.۰۱	۰.۰۰۰۸	٪۳.۲۶
انتشار وابستگی	بیشترین شارپ	HRP	حذف	٪۰.۰۱	۰.۰۰۰۹	٪۳.۲۰
انتشار وابستگی	بیشترین شارپ	HRP	عدم حذف	٪۰.۰۱	۰.۰۰۰۹	٪۳.۲۰
K-Medoids	استفاده از همه	HRP	عدم حذف	٪۰.۰۱	۰.۰۰۰۸	٪۳.۱۲
K-Means	استفاده از همه	HRP	عدم حذف	٪۰.۰۱	۰.۰۰۰۸	٪۳.۱۲
انتشار وابستگی	استفاده از همه	HRP	عدم حذف	٪۰.۰۱	۰.۰۰۰۸	٪۳.۱۲

جدول ۴-۱: نتایج مدل (ادامه)

روش خوشه‌بندی	روش انتخاب نماد در هر خوشه	روش انتخاب سبد	نقاط مغلوب	بازده	ریسک	نسبت شارپ
K-Means	بیشترین بازده	میانگین-واریانس	حذف	٪۰.۱۷	۰.۰۶۷۹	٪۳.۱۰
K-Means	بیشترین بازده	میانگین-واریانس	عدم حذف	٪۰.۱۷	۰.۰۶۷۹	٪۳.۱۰
K-Means	بیشترین شارپ	میانگین-واریانس	حذف	٪۰.۱۷	۰.۰۷۰۴	٪۳.۰۵
K-Means	بیشترین شارپ	میانگین-واریانس	عدم حذف	٪۰.۱۷	۰.۰۷۰۴	٪۳.۰۵
K-Medoids	استفاده از همه	میانگین-CVaR	عدم حذف	٪۰.۱۴	۰.۰۲۵۴	٪۲.۷۹
K-Means	استفاده از همه	میانگین-CVaR	عدم حذف	٪۰.۱۴	۰.۰۲۵۴	٪۲.۷۹
انتشار وابستگی	استفاده از همه	میانگین-CVaR	عدم حذف	٪۰.۱۴	۰.۰۲۵۴	٪۲.۷۹
انتشار وابستگی	کمترین ریسک	HRP	حذف	٪۰.۰۱	۰.۰۰۰۹	٪۲.۵۲
انتشار وابستگی	کمترین ریسک	HRP	عدم حذف	٪۰.۰۱	۰.۰۰۰۹	٪۲.۵۲
انتشار وابستگی	کمترین ریسک	میانگین-CVaR	حذف	٪۰.۲۴	۰.۰۲۸۸	٪۲.۴۶
انتشار وابستگی	کمترین ریسک	میانگین-CVaR	عدم حذف	٪۰.۲۴	۰.۰۲۸۸	٪۲.۴۶
انتشار وابستگی	بیشترین بازده	HRP	حذف	٪۰.۰۱	۰.۰۰۰۸	٪۲.۴۰
انتشار وابستگی	بیشترین بازده	HRP	عدم حذف	٪۰.۰۱	۰.۰۰۰۸	٪۲.۴۰
انتشار وابستگی	بیشترین بازده	میانگین-CVaR	حذف	٪۰.۱۵	۰.۰۴۰۴	٪۲.۲۰
انتشار وابستگی	بیشترین بازده	میانگین-CVaR	عدم حذف	٪۰.۱۵	۰.۰۴۰۴	٪۲.۲۰
K-Medoids	کمترین ریسک	HRP	حذف	٪۰.۰۰	۰.۰۰۰۸	٪۲.۱۸
K-Medoids	کمترین ریسک	HRP	عدم حذف	٪۰.۰۰	۰.۰۰۰۸	٪۲.۱۸
انتشار وابستگی	استفاده از همه	میانگین-CVaR	حذف	٪۰.۱۷	۰.۰۲۶۶	٪۱.۸۱

۴-۵- تحلیل نتایج

نتایج نهایی مدل به شکل میانگینی از بازده، ریسک و نسبت شارپ در بیست تاریخ ذکر شده ذکر بیان می‌شود. تفکیک این نتایج به این شکل است که در ابتدا برای هر یک از روش‌های خوشه‌بندی، مقایسه‌ای بین روش‌های مختلف انتخاب سبد سرمایه‌گذاری نمایش داده شود. سپس در هر روش، نتایج به روش انتخاب نماد در هر سبد نیز تفکیک می‌شوند. برای هر یک روش‌ها دو نمودار نمایش داده می‌شود که نمودار اول نتایج اجرای مدل با حذف نقاط مغلوب و نمودار دوم، بدون نتایج این حذف را نمایش می‌دهد. در نهایت، تحلیل مقایسه‌ای از دو حالت حذف و عدم حذف نقاط مغلوب در مدل ارائه خواهد شد.

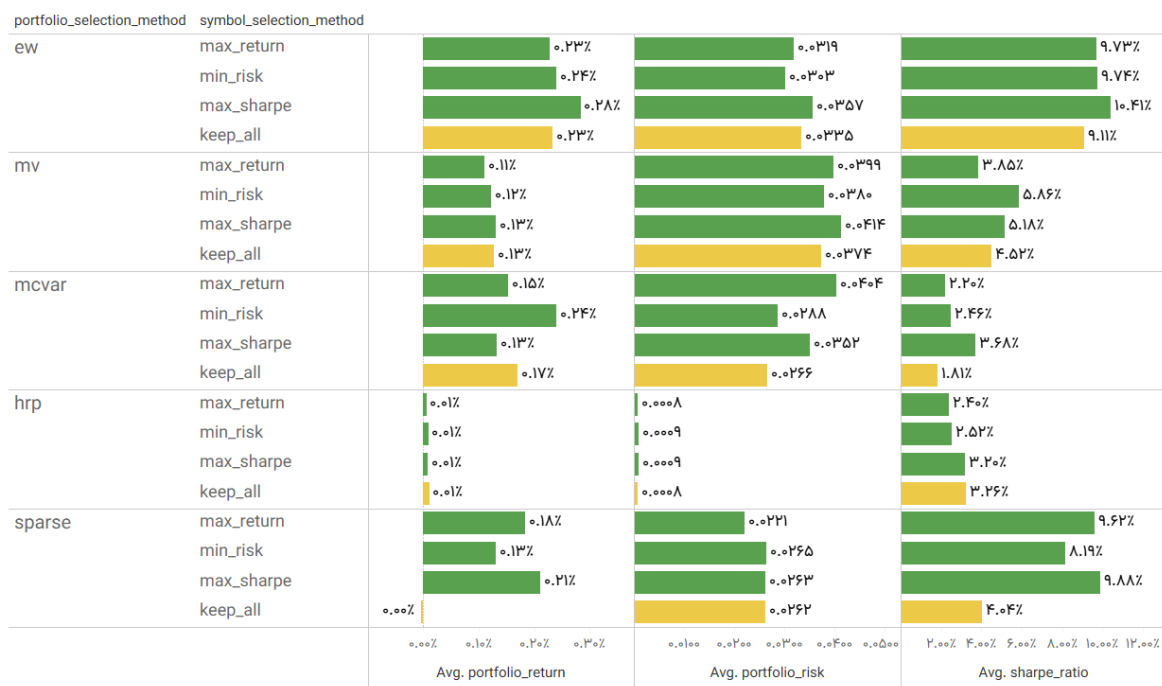
۴-۵-۱- تحلیل نتایج به تفکیک روش خوشه‌بندی

همان‌طور که در شکل‌های ۴-۵ و ۴-۶ دیده می‌شود، روش انتشار وابستگی با روش‌های انتخاب سبد Sparse و هم‌وزن بهترین نتیجه را به ما می‌دهد. در هر دوی این روش‌ها استفاده از بیشترین نسبت شارپ برای انتخاب نماد در هر خوشه بهترین نتیجه را به سرمایه‌گذار می‌دهد. همچنین در همه‌ی روش‌ها، حداقل یک استراتژی وجود دارد که در آن استفاده از خوشه‌بندی بهتر از عدم استفاده‌ی آن باشد. علاوه بر این، استفاده از روش HRP برای انتخاب سبد در این حالت، با وجود اینکه نسبت شارپ پایینی نسبت به روش‌های دیگر می‌سازد، اما ریسک بسیار پایین‌تری را نسبت به روش‌های دیگر متحمل می‌شود. بنابراین در شرایط عدم اطمینان و نوسانی بازار، می‌تواند به یکی از روش‌های جذاب برای سرمایه‌گذاران تبدیل شود. شکل‌های ۴-۷ و ۴-۸ مربوط به روش دیگر خوشه‌بندی، یعنی روش K-Means است. نتایج اجرای مدل با این روش خوشه‌بندی، روش‌های انتخاب سبد سرمایه‌گذاری HRP، Sparse، هم‌وزن، و با اختلاف کمی میانگین CVaR را پیشنهاد می‌دهد. در تمامی این روش‌ها استفاده از خوشه‌بندی بهتر از عدم استفاده از

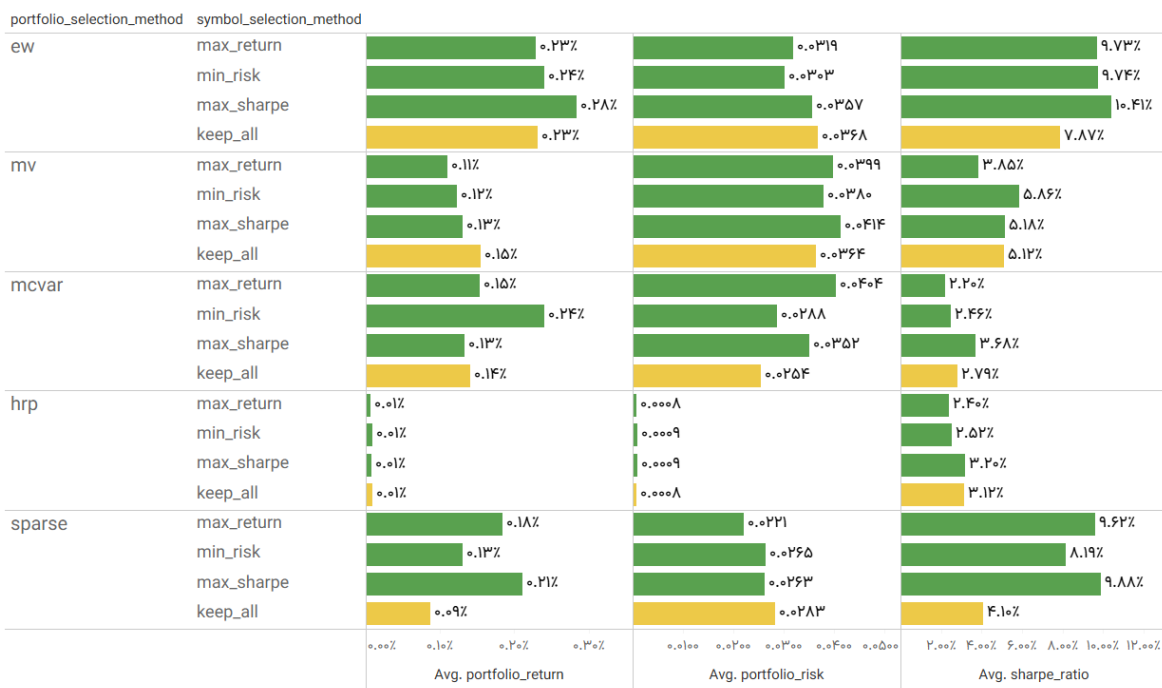
آن است. اختلاف این استفاده به خصوص در روش‌های Sparse, HRP و منیائنگین-CVaR بسیار چشم‌گیر است و عدم استفاده از خوشه‌بندی در این روش‌ها به از دست دادن سود زیادی منجر می‌شود.

سومین روش خوشه‌بندی مورد استفاده، روش K-Medoids است که نتایج استفاده از آن را می‌توان در شکل‌های ۴-۹ و ۴-۱۰ مشاهده کرد. نتایج استفاده از مدل‌های انتخاب سبد سرمایه‌گذاری در این روش، تا حد زیادی مشابه است و اختلاف چشم‌گیری در یک روش خاص دیده نمی‌شود. به جز در روش هم‌وزن، در باقی روش‌های انتخاب سبد استفاده از خوشه‌بندی نتیجه‌ی بهتری را عاید سرمایه‌گذار می‌کند.

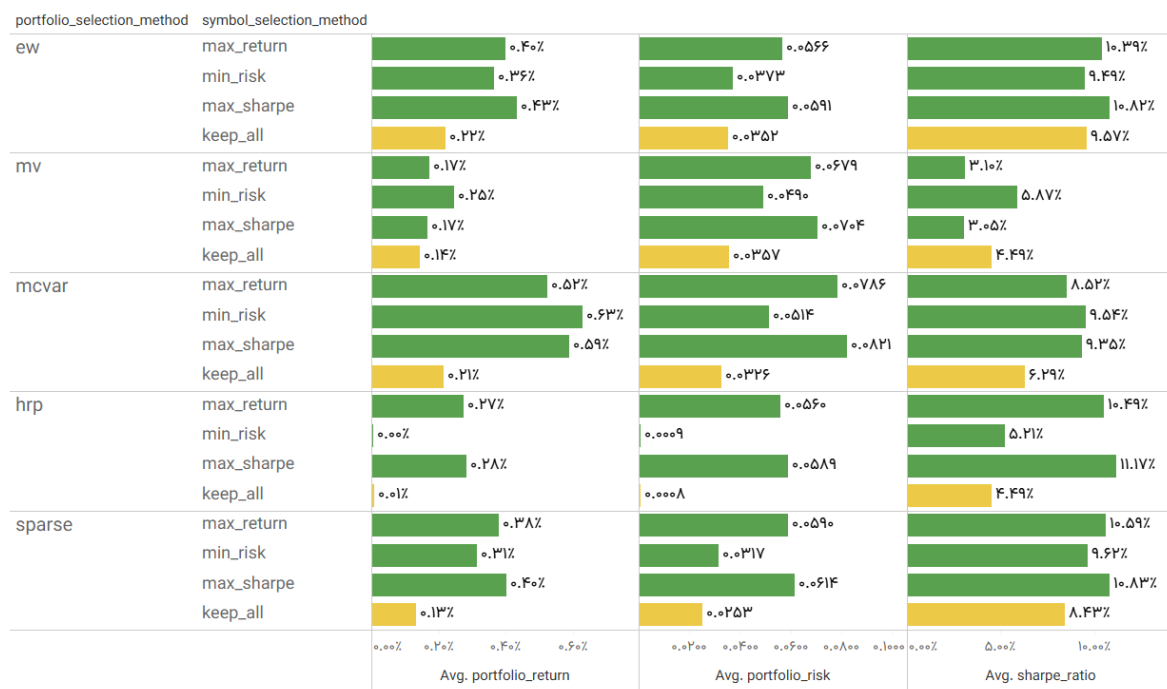
به طور کلی، بهترین نتایج در روش K-Means به دست می‌آید که در بسیاری از موارد نسبت شارپ بیشتر از ده درصد را رقم می‌زند. بهترین حالت سرمایه‌گذاری مطابق مدل پیشنهادی، استفاده از روش HRP برای انتخاب سبد سرمایه‌گذاری و استفاده از روش بیشترین نسبت شارپ، برای انتخاب نماد در هر خوشه است که منجر به نسبت شارپ بیش از ۱۱ درصد می‌شود. این در حالی است که استفاده از روش K-Medoids، در بیشترین حالت نسبت شارپ ۸ درصدی خواهد داشت.



شکل ۴-۵: نتایج خوشه‌بندی انتشار وابستگی با حذف نقاط مغلوب



شکل ۴-۶: نتایج خوشه‌بندی انتشار وابستگی بدون حذف نقاط مغلوب



شکل ۴-۷: نتایج خوشه‌بندی K-Means با حذف نقاط مغلوب



شکل ۴-۸: نتایج خوشه‌بندی K-Means بدون حذف نقاط مغلوب



شکل ۴-۹: نتایج خوشه‌بندی K-Medoids با حذف نقاط مغلوب



شکل ۴-۱۰: نتایج خوشه‌بندی K-Medoids بدون حذف نقاط مغلوب

۴-۵-۲- تحلیل نتایج به تفکیک روش انتخاب سبد سرمایه‌گذاری

تفکیک دیگری که برای بررسی نتایج مدل می‌توان داشت، تفکیک بر اساس روش انتخاب سبد سرمایه‌گذاری است. به این شکل می‌توان روش‌های مختلف خوشه‌بندی را در هر یک روش‌های انتخاب سبد با یکدیگر مقایسه کرد. نحوه‌ی نمایش نتایج مانند بخش قبلی است و تنها نوع تفکیک نتایج متفاوت شده است.

شکل‌های ۴-۱۱ و ۴-۱۲، مقایسه‌ای از روش‌های موجود در روش انتخاب سبد هم‌وزن را نشان می‌دهد. به طور کلی، روش‌های K-Means و انتشار وابستگی عملکرد بسیار مناسبی در خوشه‌بندی داشته‌اند. بهترین روش با انتخاب بیشترین نسبت شارپ در خوشه‌های K-Means به دست می‌آید. نتایج نشان می‌دهد که در استفاده از روش هم‌وزن، روش K-Medoids بر خلاف دو روش دیگر اثربخشی پایین‌تری از عدم استفاده از خوشه‌بندی داشته است. با این حال در تمام روش‌های ذکرشده، حذف نقاط مغلوب نتیجه‌ی بهتری از عدم حذف نقاط مغلوب داشته است.

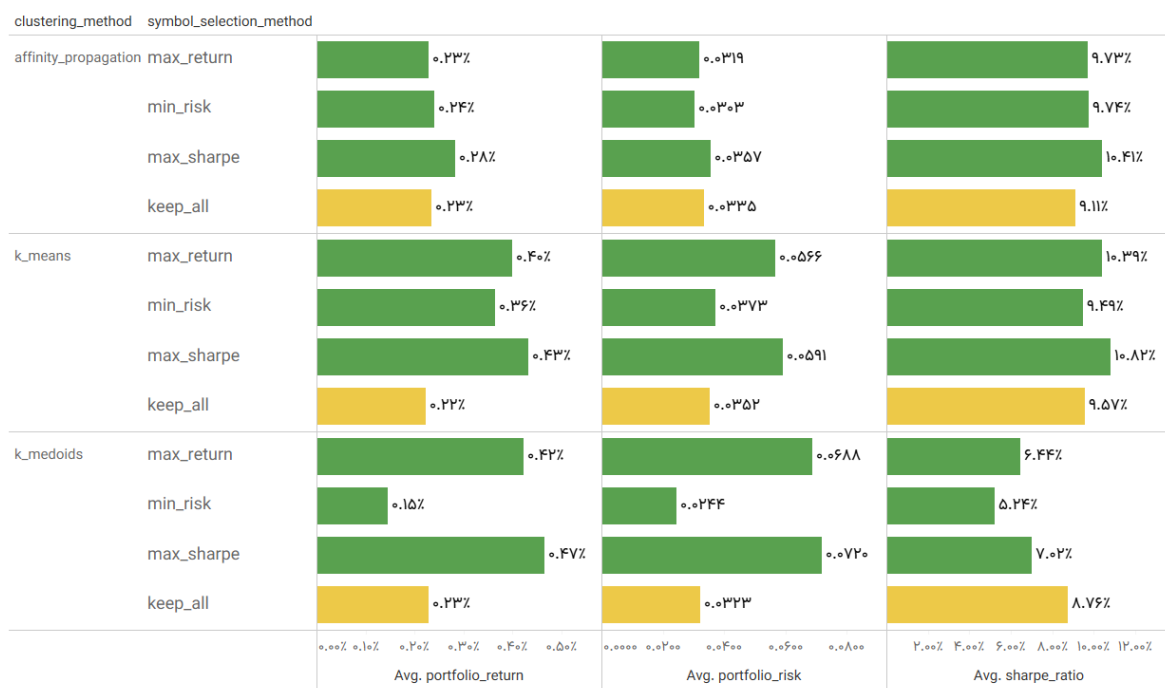
نتایج روش میانگین-واریانس در شکل‌های ۴-۱۳ و ۴-۱۴ نمایش داده شده است. در این روش بر خلاف روش هم‌وزن، بهترین نتیجه با انتخاب نماد دارای کمترین ریسک در خوشه‌های K-Medoids به دست آمده است. همچنین تنها در این روش خوشه‌بندی است که با انتخاب همه‌ی نمادهای هر خوشه، حذف نقاط مغلوب نتیجه‌ی بهتری از عدم حذف آن‌ها را به دست می‌دهد.

استفاده از شاخص ریسک CVaR به جای واریانس در مدل ماکوویتز، نتایج بهتری ارائه می‌دهد که در شکل‌های ۴-۱۵ و ۴-۱۶ مشاهده می‌شود. بهترین استراتژی در این روش استفاده از خوشه‌بندی K-Means است که اختلاف معناداری با سایر روش‌های خوشه‌بندی دارد. همچنین به طور کلی حداقل یک استراتژی در هر روش وجود دارد که خوشه‌بندی نمادها را توجیه کند.

روش HRP که نتایج آن در شکل‌های ۴-۱۷ و ۴-۱۸ وجود دارد نیز همچون روش قبل، روش K-Means را برای خوشه‌بندی پیشنهاد می‌کند. با این انتخاب، استفاده از بیشترین نسبت شارپ در هر خوشه بهترین نتیجه را عاید سرمایه‌گذار می‌کند. با این حال روش انتشار وابستگی عملکرد ضعیفی از خود نمایش می‌دهد. همچنین مشاهده می‌شود که حذف نقاط مغلوب در هر خوشه نتیجه‌ی بهتری را از عدم حذف این نقاط در هر خوشه به دست می‌دهد. به شکل کلی روش HRP با متعادل‌سازی ریسک خطر بسیار پایین‌تری را نسبت به سایر روش‌ها متوجه سرمایه‌گذار می‌کند.

روش Sparse برای انتخاب سبد سرمایه‌گذاری، عملکرد بسیار مناسبی از خود به نمایش می‌گذارد که در شکل‌های ۴-۱۹ و ۴-۲۰ مشاهده می‌شود. بهترین نتایج این مدل با استفاده از روش K-Means و پس از آن با استفاده از روش انتشار وابستگی ایجاد می‌شود. استفاده از خوشه‌بندی در انتخاب سبد به روش Sparse نتایج بسیار بهتری نسبت به عدم استفاده از آن را ایجاد می‌کند که در تمام روش‌های خوشه‌بندی قابل قبول و مناسب است.

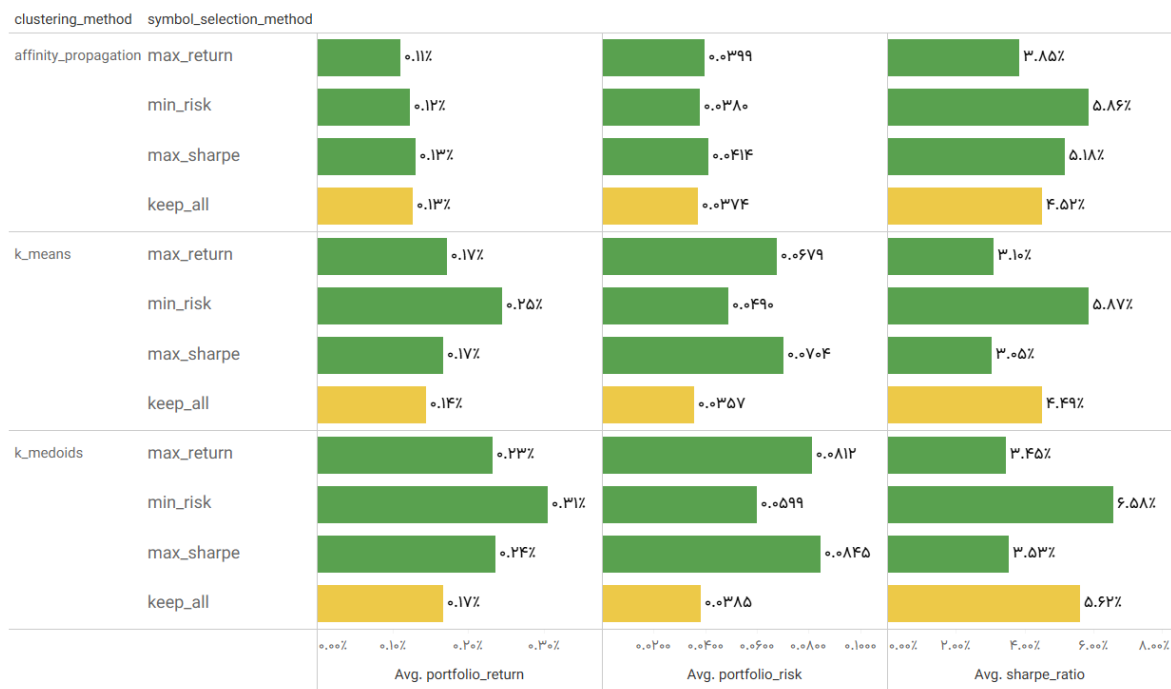
به طور کلی نتایج نهایی نشان می‌دهد که روش Sparse نتایج بهتری نسبت به سایر روش‌های انتخاب سبد سرمایه‌گذاری ایجاد می‌کند. با این حال بهترین نتیجه در روش HRP و با خوشه‌بندی K-Means مشاهده می‌شود. در روش HRP، استفاده از خوشه‌بندی نتایج بسیار مناسب‌تری از عدم استفاده از آن را عاید سرمایه‌گذار می‌کند.



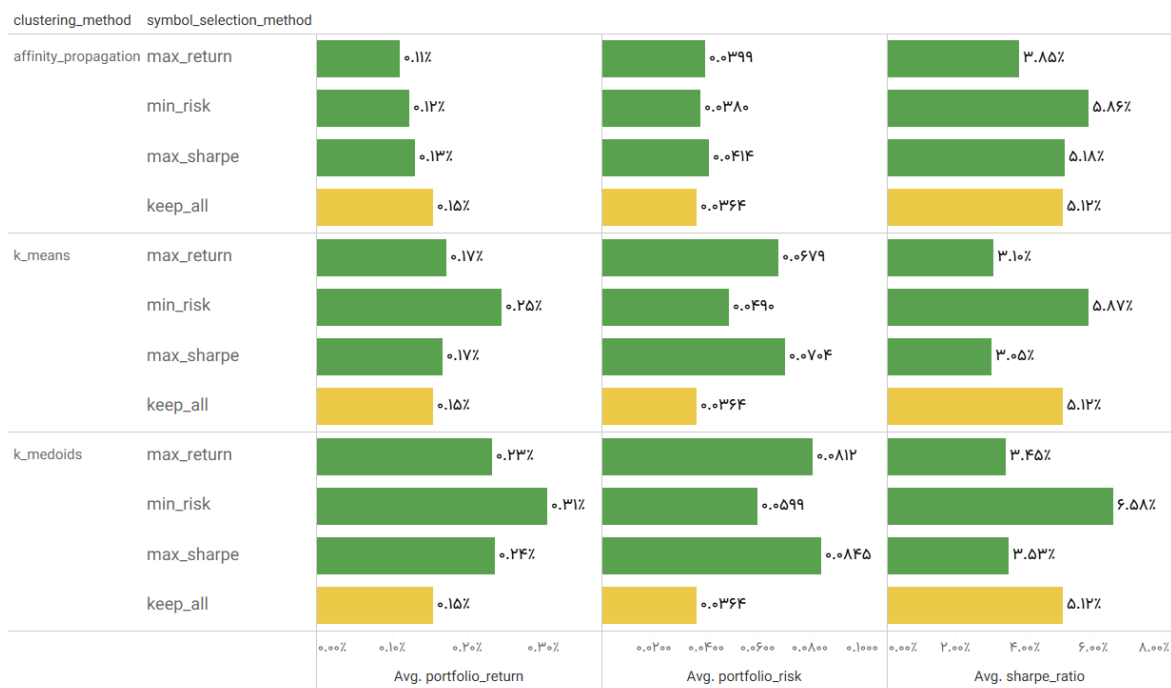
شکل ۴-۱۱: نتایج روش هموزن با حذف نقاط مغلوب



شکل ۴-۱۲: نتایج روش هموزن بدون حذف نقاط مغلوب



شکل ۴-۱۳: نتایج روش میانگین-واریانس با حذف نقاط مغلوب



شکل ۴-۱۴: نتایج روش میانگین-واریانس بدون حذف نقاط مغلوب



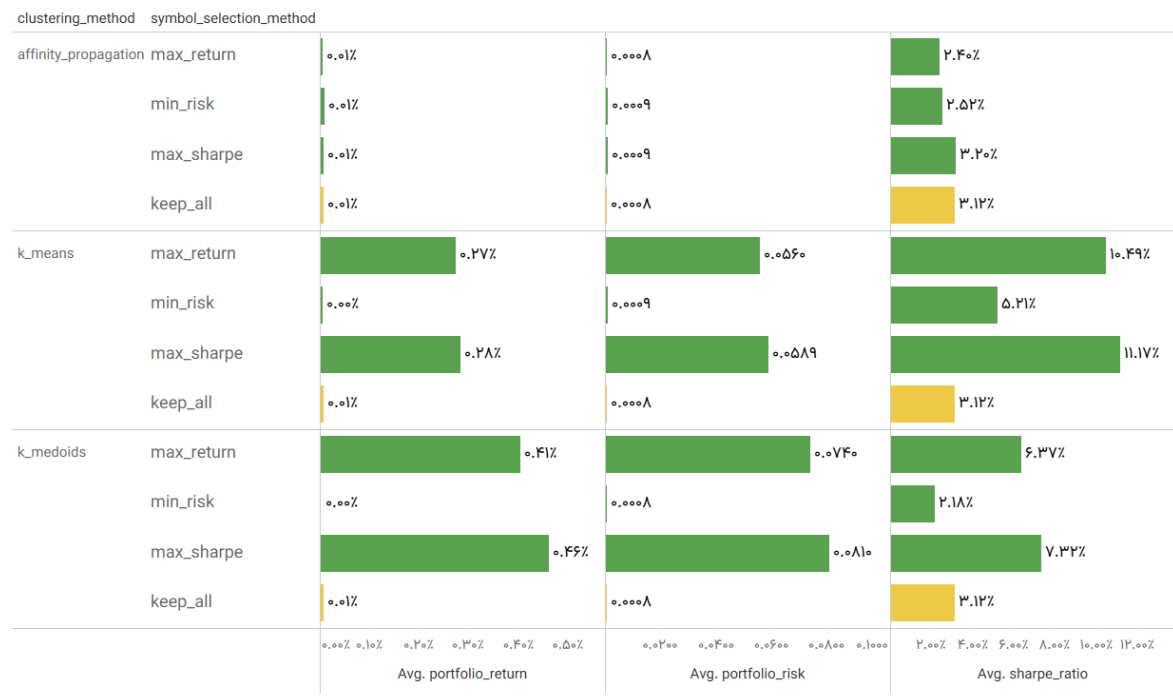
شکل ۴-۱۵: نتایج روش میانگین-CVaR با حذف نقاط مغلوب



شکل ۴-۱۶: نتایج روش میانگین-CVaR بدون حذف نقاط مغلوب



شکل ۴-۱۷: نتایج روش HRP با حذف نقاط مغلوب



شکل ۴-۱۸: نتایج روش HRP بدون حذف نقاط مغلوب



شکل ۴-۱۹: نتایج روش Sparse با حذف نقاط مغلوب

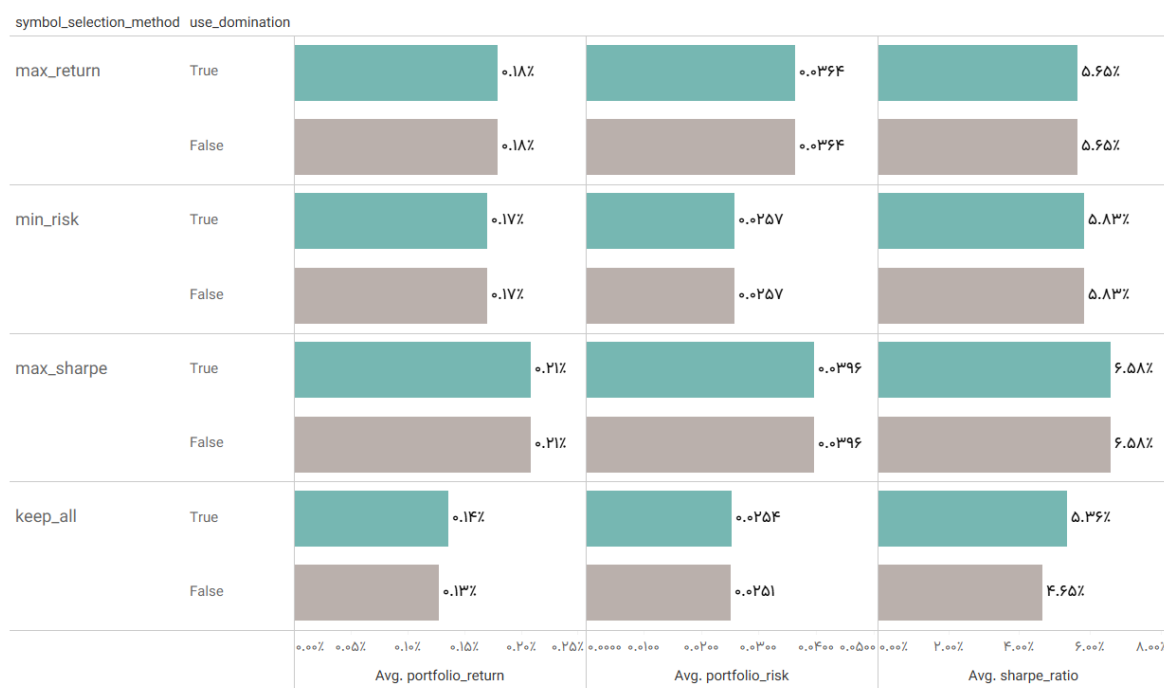


شکل ۴-۲۰: نتایج روش Sparse بدون حذف نقاط مغلوب

۴-۵-۳- تحلیل نتایج حذف یا عدم حذف نقاط مغلوب

شکل ۴-۲۱، نتایج کلی مربوط به حذف نقاط مغلوب در روش‌های مختلف را نشان می‌دهد. تفکیک نمودارها بر اساس روش انتخاب نماد در هر خوشه انجام شده است. همان طور که مشاهده می‌شود، در هیچ یک از سه روش انتخاب یک نماد در هر خوشه، یعنی روش‌های بیشترین بازده، کمترین ریسک و بیشترین نسبت شارپ، تفاوتی از نظر حذف یا عدم حذف نقاط مغلوب در هر خوشه نمی‌کند.

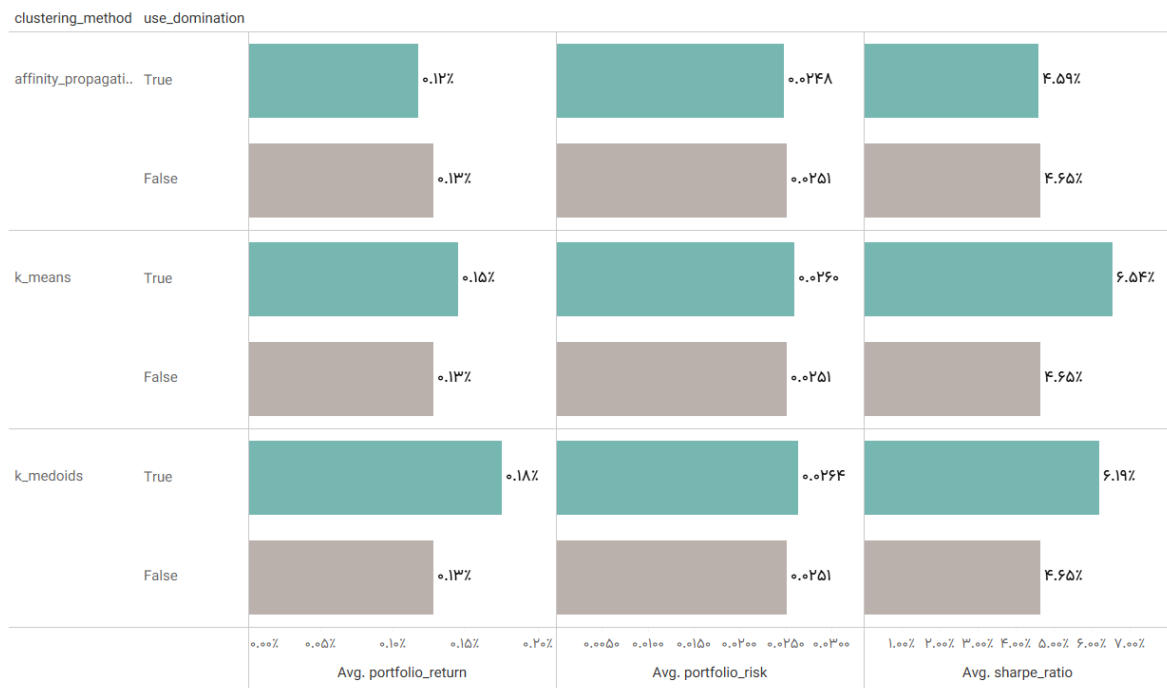
این مسئله به این معناست که در همه‌ی روش‌ها حتی با وجود عدم حذف نقاط مغلوب، نمادهایی در هر خوشه انتخاب شده‌اند که جزو این نقاط نبوده‌اند. شرایط اما در نگهداری همه‌ی نمادها در هر خوشه متفاوت است. در این حالت، بهترین عملکرد از نظر نسبت شارپ متوجه روش‌هایی است که در آن‌ها نمادهای مغلوب از خوشه‌ها حذف شده‌اند.



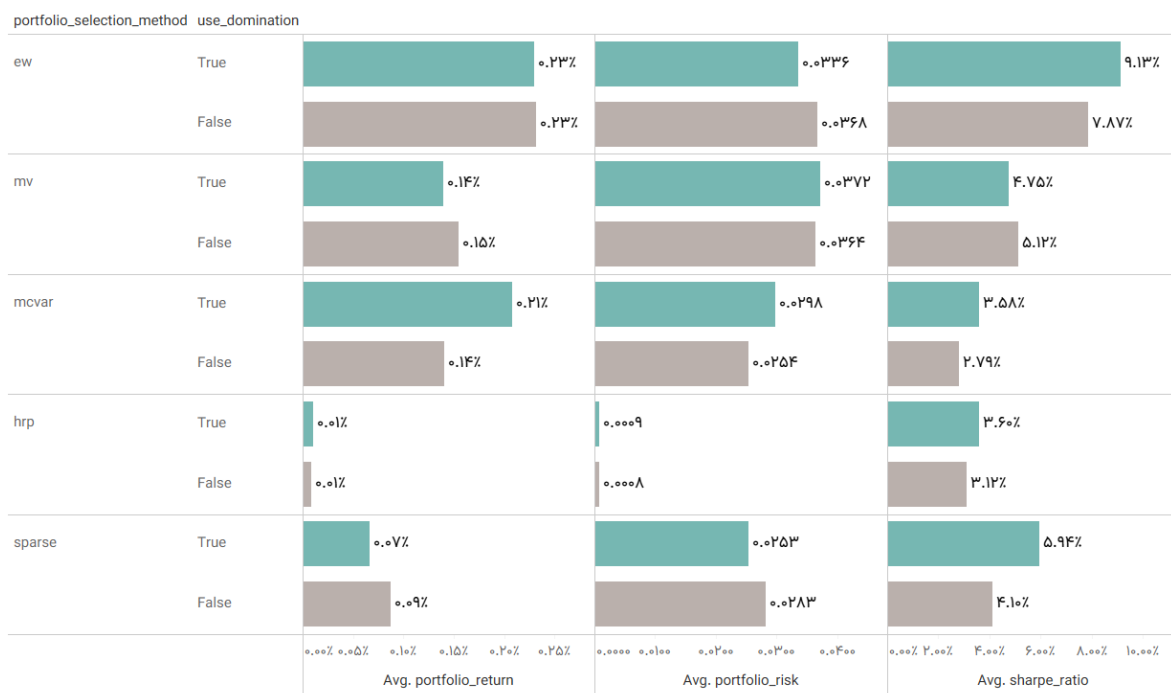
شکل ۴-۲۱: نتایج مربوط به حذف نقاط مغلوب

همان طور که مشاهده شد، به طور کلی و بر اساس میانگین نسبت شارپ، عملکرد حذف نمادهای مغلوب بهتر از عدم حذف این نقاط بوده است. برای نتیجه‌گیری بهتر، جزئیات حالت نگهداری همه‌ی نمادها در خوشه‌ها، در شکل‌های ۴-۲۲ و ۴-۲۳ مشاهده می‌شود. نمودار اول، مربوط به تفکیک بر اساس روش خوشه‌بندی است. این نمودار نشان می‌دهد که حذف نمادهای مغلوب در روش‌های خوشه‌بندی K-Means و K-Medoids عملکرد بسیار بهتری ایجاد می‌کند. این در حالی است که در روش انتخاب وابستگی، عملکرد دو روش تقریباً با هم برابر است.

نمودار دوم مقایسه‌ی روش‌های مختلف انتخاب سبد سرمایه‌گذاری را در حالت عدم انتخاب یک نماینده برای هر خوشه را نمایش می‌دهد. این نمودار نیز نشان‌دهنده‌ی عملکرد کلی مناسب حذف نقاط مغلوب در خوشه‌ها است. روش میانگین-واریانس تنها روشی است که حذف نقاط مغلوب موجب کاهش نسبت شارپ سبد سرمایه‌گذاری نهایی شده است.



شکل ۴-۲۲: نتایج مربوط به نگهداری همه‌ی نمادها به تفکیک روش خوشه‌بندی



شکل ۴-۲۳: نتایج مربوط به نگهداری همه‌ی نمادها به تفکیک روش انتخاب سبد سرمایه‌گذاری

۴-۶- جمع‌بندی

در این فصل به بررسی پارامترهای ورودی، اجرای مدل و تحلیل نتایج حاصل‌شده پرداخته شد. این مدل از داده‌های قیمتی رمزارزها در سال‌های اخیر استفاده می‌کند و در هر یک از مراحل، از روش‌های گوناگونی بهره می‌گیرد. این روش‌های گوناگون به ما کمک می‌کند که ارتباط بین روش‌های خوشه‌بندی و روش‌های انتخاب سرمایه‌گذاری در بخش تحلیل نتایج برای سرمایه‌گذار مشخص شود. نمودارهای حاصل‌شده در بخش تحلیل نتایج، به سرمایه‌گذاران کمک می‌کند با توجه به میزان ریسک‌پذیری خود بهترین روش را برای سرمایه‌گذاری با استفاده از مدل حاضر اتخاذ کنند و در سرمایه‌گذاری خود در بازار رمزارزها به کار برند.

فصل پنجم

نتیجه‌گیری و پیشنهادات برای تحقیقات آتی

۵-۱- مقدمه

این فصل در ابتدا به جمع‌بندی پژوهش حاضر و بیان خلاصه‌ای از مدل و نتایج حاصل‌شده از آن می‌پردازد و در ادامه پیشنهاداتی را جهت توسعه‌ی این مدل برای پژوهشگران آتی ارائه می‌کند تا با استفاده از این پیشنهادات، قابلیت توسعه‌ی مدل ذکرشده وجود داشته باشد.

۵-۲- جمع‌بندی و خلاصه‌ای از پژوهش

سرمایه‌گذاران به طور کلی با انجام سرمایه‌گذاری در هر یک از بازارهای موجود، در معرض انواع ریسک‌های سیستماتیک و غیرسیستماتیک قرار می‌گیرند. ریسک سرمایه‌گذاری به خصوص در بازار نوظهور و کمترشناخته‌شده‌ای مثل بازار رمزارزها دوجندان می‌گردد. پژوهش حاضر با هدف کاهش ریسک سرمایه‌گذاری با انتخاب یک سبد سرمایه‌گذاری به شکل بهینه ارائه شده است.

مدل استفاده‌شده در این پژوهش، از مراحل مختلفی تشکیل شده است. سرمایه‌گذار در تمام مراحل قادر است تا پارامترهای ورودی مدل را بر حسب نیاز خود تغییر دهد یا در تاریخ‌های خاصی و بر اساس افق سرمایه‌گذاری خود مدل را اجرا کند. این مدل تلاش می‌کند در هر مرحله، از شیوه‌های نوینی استفاده کند که در مطالعات اخیر مورد آزمون قرار گرفته و نتیجه‌بخش بوده اند. به همین منظور، پس از دریافت داده‌های اولیه، تمیزسازی آن‌ها و تقسیم آن‌ها به داده‌های گذشته (برای اجرای مدل و به دست آوردن نسبت هر

نماد در سبد سرمایه‌گذاری) و داده‌های آینده (برای آزمون و سنجش عملکرد مدل)، نسبت به خوشه‌بندی نمادها بر اساس اطلاعات قیمتی اقدام می‌شود. این خوشه‌بندی نمادها می‌تواند با استفاده از یکی از روش‌های ارائه‌شده، یعنی انتشار وابستگی، K-Means و K-Medoids صورت پذیرد.

پس از مشخص شدن خوشه‌ی هر نماد، یک یا تعدادی از نمادها در هر گروه انتخاب می‌شود. این انتخاب می‌تواند از طریق حذف نمادهای مغلوب از نظر داشتن بازده کمتر و ریسک بیشتر، و نیز از طریق انتخاب نماد دارای بیشترین بازده، کمترین ریسک یا بیشترین نسبت شارپ صورت گیرد. بعد از مشخص شدن نماینده یا نمایندگان خوشه‌ها، تمام نمادهای باقی‌مانده به مرحله‌ی انتخاب سبد سرمایه‌گذاری منتقل می‌شوند که می‌تواند بر اساس یکی از روش‌های ارائه‌شده یعنی روش هم‌وزن، روش میانگین-واریانس، روش میانگین-CVaR، روش HRP و یا روش Sparse انجام شود. در نهایت عملکرد مدل بر اساس معیارهایی مانند بازده، ریسک و نسبت شارپ محاسبه و اعلام می‌شود.

با تعیین پارامترهای اولیه و اجرای مدل در شرایط مختلف، عملکرد روش‌های گوناگون استفاده‌شده در هر مرحله به دست می‌آید. نتایج نهایی مدل نشان می‌دهد که در اکثر موارد به خصوص در روش Sparse که یکی از بهترین نتایج را در مرحله‌ی بهینه‌سازی سبد سرمایه‌گذاری رقم می‌زند، استفاده از خوشه‌بندی موجب بهبود چشم‌گیر نتایج نهایی می‌شود. از بین روش‌های خوشه‌بندی، روش K-Means عملکرد نهایی بهتری را از نظر شاخص نسبت شارپ نسبت به سایر روش‌ها رقم می‌زند. همچنین، حذف نقاط مغلوب از سبد سرمایه‌گذاری پیش از مرحله‌ی انتخاب سبد، موجب حاصل شدن نتایج بهتری نسبت به عدم حذف نقاط مغلوب می‌گردد.

در نهایت مدلی که بهترین نتیجه را از نظر بیشینه نسبت شارپ به دست می‌دهد، مدلی است که در مرحله‌ی خوشه‌بندی از روش K-Means، در مرحله انتخاب نماد در هر خوشه از روش انتخاب نماد دارای بیشترین نسبت شارپ، و در مرحله‌ی انتخاب سبد سرمایه‌گذاری از روش HRP استفاده می‌کند. همچنین سبد سرمایه‌گذاری که منجر به حاصل شدن بیشترین بازده سرمایه‌گذاری می‌شود متعلق به مدلی است که

از روش K-Means برای خوشه‌بندی، از روش کمترین ریسک برای انتخاب نماد و از روش میانگین-CVaR برای انتخاب سبد استفاده می‌کند. از نظر سبدهای دارای کمترین ریسک نیز مدل‌های استفاده‌کننده از روش HRP به عنوان روش انتخاب سبد سرمایه‌گذاری، به شکل چشم‌گیری پیش‌تاز هستند.

۵-۳- پیشنهادات برای تحقیقات آتی

پژوهش حاضر دارای قابلیت بسط و بهبود نتایج با استفاده از نوآوری‌هایی است که پژوهشگران آتی می‌توانند به آن بیفزایند. از جمله پیشنهاداتی که می‌توان برای تحقیقات آینده از آن‌ها بهره برد موارد زیر به طور خلاصه ذکر می‌گردد:

۱- اجرا و تحلیل نتایج مدل در بازارهای دیگر مالی مانند بورس اوراق بهادار تهران، بورس کشورهای خارجی و بازار تبادل ارزهای خارجی (فارکس).

۲- اجرا و تحلیل نتایج مدل با استفاده از ترکیبی از دارایی مختلف مانند رمزارز، سهام، نفت و انرژی، فلزات گرانبها و ارزهای خارجی.

۳- اجرا و تحلیل نتایج مدل با استفاده از ترکیبی از صندوق‌های سرمایه‌گذاری.

۴- اجرا و تحلیل نتایج مدل در شرایط مجاز بودن فروش استقراضی در انتخاب سبد سرمایه‌گذاری.

۵- استفاده از روش‌های دیگر خوشه‌بندی مانند خوشه‌بندی مخفی مارکوف، خوشه‌بندی ترکیبی گاوسی، خوشه‌بندی DBSCAN و خوشه‌بندی OPTICS.

۶- استفاده روش‌های انتخاب چند نماد در هر خوشه، مانند انتخاب جفت-نمادهای دارای بیشترین بازده و کمترین ریسک در هر خوشه.

۷- استفاده از روش‌های دیگر انتخاب سبد سرمایه‌گذاری، مانند روش انتخاب سبد بازار و روش‌های تغییر سنجی ریسک در مدل مارکوویتز.

۸- استفاده از قیمت‌های پیش‌بینی‌شده توسط مدل‌های رگرسیونی به جای استفاده از قیمت‌های تاریخی در اجرای مدل.

۹- استفاده از معیارهای دیگر ارزیابی عملکرد سبد سرمایه‌گذاری مانند نسبت ترینر.

۱۰- استفاده از شاخص‌های تحلیل تکنیکال مانند RSI، MACD، EMA و... به عنوان پارامترهای ورودی خوشه‌بندی در کنار قیمت.

فهرست مراجع

- آقامحمدی، احمد؛ اوحدی، فریدون؛ صیقلی، محسن؛ بنی‌مهد، بهمن. ۱۳۹۹، برآورد ریسک سرمایه‌گذاری در یک پرتفوی ارز دیجیتال و بهینه‌سازی آن با استفاده از روش ارزش در معرض خطر. *دانش مالی تحلیل اوراق بهادار*، ۱۳(۴۷)، ص ۱۷-۳۱.
- ابونوری، اسماعیل؛ تهرانی، رضا؛ شامانی، مسعود. ۱۳۹۷، عملکرد پرتفولیوهای مبتنی بر ریسک تحت شرایط مختلف در بازار سهام. *فصلنامه‌ی اقتصاد مالی*، ۱۲(۴۵)، ص ۵۱-۷۱.
- امیری، مقصود؛ محبوب قدسی، مهسا. ۱۳۹۴، مدل برنامه‌ریزی خطی فازی برای مسئله‌ی انتخاب سبد سهام بهینه. *مجله‌ی مهندسی مالی و مدیریت اوراق بهادار*، ۲۳، ص ۱۰۵-۱۱۸.
- پیکانی، پژمان؛ روغنیان، عماد. ۱۳۹۴، به‌کارگیری تحلیل پوششی داده‌ها و بهینه‌سازی استوار در مسئله‌ی انتخاب سبد سرمایه. *مجله‌ی تحقیق در عملیات در کاربردهای آن*، ۱۲(۱)، ص ۶۱-۷۸.
- راعی، رضا؛ و تلنگی، احمد. ۱۳۸۳، مدیریت سرمایه‌گذاری پیشرفته. سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها، تهران.
- زمردیان، غلامرضا؛ آزاد، محمد؛ رجب‌زاده، محمدرضا. ۱۳۹۸، مقایسه‌ی توان پیش‌بینی سنج‌های ریسک اوراق با درآمد ثابت در تعیین قیمت‌ها. *راهبرد مدیریت مالی*، ۷(۲۶)، ص ۱۷۵-۱۹۹.
- شهرستانی، حمید؛ بیدآباد، بیژن؛ ثوابی اصل، فرهاد. ۱۳۸۹، توسعه‌ی نظریه‌ی مارکوویتز-شارپ و مرز کارای جدید. *فصلنامه‌ی پژوهش‌های اقتصادی*، ۱۰(۲)، ص ۴۳-۶۰.
- صباحی، سوده؛ مخاطب‌رفیعی، فریمه؛ و رستگار، محمدعلی. ۱۳۹۹، بهینه‌سازی سبد سرمایه‌گذاری با دارایی‌های متنوع. *اقتصاد پولی مالی*، ۲۷(۱۹)، ص ۲۴۹-۲۷۸.

عباس‌نژاد، علی‌اکبر. ۱۳۸۰، ارزیابی مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران بر اساس فرآیند تحلیل سلسله‌مراتبی، پایان‌نامه‌ی کارشناسی ارشد، دانشگاه امام صادق.

قره خانی، محسن؛ سجادی، سیدجعفر؛ صفری، احرام. ۱۳۹۲، بهینه‌سازی استوار سبد مالی با رویکرد CAPM. مدیریت تولید و عملیات، ۴(۱۱) (پیاپی ۶)، ۶۱-۶۸.

قهطرانی، علیرضا. ۱۳۹۱، به‌کارگیری بهینه‌سازی استوار در مسئله‌ی انتخاب سبد سرمایه. پایان‌نامه‌ی کارشناسی ارشد مهندسی مالی، دانشگاه خواجه نصیرالدین طوسی.

کیانی هرچگانی، مائده؛ نبوی چاشمی، سید علی؛ معماریان، عرفان. ۱۳۹۳، بهینه‌سازی سبدسهم براساس حداقل سطح پذیرش ریسک کل و اجزای آن با استفاده از روش الگوریتم ژنتیک. فصل‌نامه‌ی علمی پژوهشی دانش سرمایه‌گذاری، ۳(۱۱)، ص ۱۲۵-۱۶۴.

میرلوحی، سیدمجتبی؛ محمدی تودشکی، نیما. ۱۳۹۹، تشکیل سبد سرمایه‌گذاری بهینه در بورس اوراق بهادار تهران با استفاده از روش‌های خوشه‌بندی سلسله‌مراتبی و تفکیکی. فصلنامه علمی پژوهشی دانش سرمایه‌گذاری، ۹(۳۴)، ص ۳۳۳-۳۵۴.

میرمحمدی، سید ابراهیم؛ معدنچی زاج، مهدی؛ پناهیان، حسین؛ جباری، حسین. ۱۴۰۱، انتخاب سبد سهام با رویکرد ترکیبی تشابه ریسک و تحلیل عاملی بر پایه تغییر رژیم مارکوف. نشریه تصمیم‌گیری و تحقیق در عملیات، ۷(۱)، ص ۱۲۹-۱۴۲.

Arévalo, A., León, D., Hernandez, G., 2019. Portfolio Selection Based on Hierarchical Clustering and Inverse-Variance Weighting. Computational Science. ICCS 2019. Lecture Notes in Computer Science, vol 11538. Springer, Cham.

Baser, P., & Saini, J., 2015. Agent based Stock Clustering for Efficient Portfolio Management. International Journal of Computer Applications, 116, 35-41.

Ben-Tal, A., Nemirovski, A., 2000. Robust solutions of linear programming problems contaminated with uncertain data. Mathematical programming, 88(3), pp. 411-424.

- Bertsimas, D., Sim, M., 2003. Robust discrete optimization and network flows. *Mathematical Programming Series*, 98, pp. 49-71.
- Brauneis, A., Mestel, R., 2019. Cryptocurrency-portfolios in a mean-variance framework. *Finance Research Letters* (2021), pp. 259-264.
- Brière, M., Oosterlinck, K., Szafarz, A., 2015. Virtual currency, tangible return: Portfolio diversification with bitcoin. *J. Asset Manag. Journal of Asset Management*, 16, pp. 365-373.
- Brodie, J., Daubechies, I., De Mol, C., Giannone, D., & Loris, I. (2009). "Sparse and stable Markowitz portfolios." *Proceedings of the National Academy of Sciences*, 106(30), 12267-12272.
- Chang, C., Lin, Z., Koc, W., Chou, C., Huang, S., 2016. Affinity Propagation Clustering for Intelligent Portfolio Diversification and Investment Risk Reduction. *International Conference on Cloud Computing and Big Data (CCBD)*, Macau, China, 2016 pp. 145-150.
- Chowdhury, S., Schoen, M. P., 2020. Research Paper Classification using Supervised Machine Learning Techniques. *Intermountain Engineering, Technology and Computing (IETC)*, Orem, UT, USA, 2020, pp. 1-6.
- Čuljak, M., Tomić, B., Žiković, S., 2022. Benefits of sectoral cryptocurrency portfolio optimization. *Research in International Business and Finance*, 60 (2022) 101615.
- Das, J. D., Bowala, S., Thulasiram R. K., Thavaneswaran, A., 2023. Portfolio Diversification with Clustering Techniques. *IEEE Symposium Series on Computational Intelligence (SSCI)*, Mexico City, Mexico, pp. 97-102.
- Datta, B.N., Sahoo, B. (2021). *Machine Learning, Regression and Optimization*. In: Sinha, B.K., Mollah, M.N.H. (eds) *Data Science and SDGs*. Springer, Singapore.
- Demiralay, S., Bayracı, S., 2021. Should stock investors include cryptocurrencies in their portfolios after all? Evidence from a conditional diversification benefits measure. *International Journal of Finance & Economics*, 26, pp. 6188-6204.
- Ding, Y., Zhang, B., 2009. Optimal portfolio of safety-first models. *Journal of Statistical Planning and Inference*, 139(2009), pp. 2952-2962.
- Estrada, J. 2007. Mean-semivariance behavior: Downside risk and capital asset pricing. *International Review of Economics and Finance*, 16(2), pp. 169-185.

- Goh, Y.C., Cai, X.Q., Theseira, W., 2020. Evaluating human versus machine learning performance in classifying research abstracts. *Scientometrics* 125, 1197–1212 (2020).
- Guesmi, K., Saadi, S., Abid, I., Ftiti, Z., 2019. Portfolio diversification with virtual currency: Evidence from bitcoin. *International Review of Financial Analysis*, 63, pp. 431-437.
- Kadam, V. S., Kanhere, S., Mahindrakar, S. 2020. Regression techniques in machine learning & applications: A review. *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)* (10):826–830.
- Kataoka, S., 1963. A Stochastic Programming Model. *The Econometric Society*, 31(1/2), pp. 181-196.
- Kaufmann, L., Rousseeuw, P., 1987. Clustering by Means of Medoids. *Data Analysis based on the L1-Norm and Related Methods*. 405-416.
- Li, R., Liu, M., Xu, D., Gao, J., Wu, F., Zhu, L., 2022. A Review of Machine Learning Algorithms for Text Classification. In: Lu, W., Zhang, Y., Wen, W., Yan, H., Li, C. (eds) *Cyber Security. CNCERT 2021. Communications in Computer and Information Science*, vol 1506. Springer, Singapore.
- Liu, W., 2019. Portfolio diversification across cryptocurrencies. *Finance Research Letters*, 29, pp. 200-205.
- Markowitz, H. M., 1952. Portfolio Selection, *The Journal of Finance*, 7(1), pp. 77-91.
- Markowitz, H. M., 1959. *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, New York.
- Mba, J. C., Pindza, E., Koumba U., 2018. A differential evolution copula-based approach for a multi-period cryptocurrency portfolio optimization. *Financial Markets and Portfolio Management*, 32, pp. 399-418.
- Nedaie, Ali & Najafi, Abbas. (2015). Polar support vector machine: Single and multiple outputs. *Neurocomputing*. 171. 10.1016/j.neucom.2015.06.029.
- Nourahmadi, M., Sadeqi, H., 2023. Portfolio Diversification Based on Clustering Analysis. *Iranian Journal of Accounting, Auditing and Finance*, 7(3), 1-16.

Orti F. J., Saez J., Terceno A., 2002. On The Treatment of Uncertainty in Portfolio Selection. *Fuzzy Economic Review*, 7(2), pp. 59-80.

Panda, K., 2023. Analysis of Optimal Portfolio Management Using Hierarchical Clustering.

Platanakis, E., Sutcliffe, C., Urquhart, A., 2018. Optimal vs naïve diversification in cryptocurrencies. *Economics Letters*, 171, 93–96.

Platanakis, E., Urquhart, A., 2019. Portfolio management with cryptocurrencies: The role of estimation risk. *Economics Letters*, 177, pp. 76-80.

Redmond, P. L., Trono, J. A., Kronenberg, D., 2011. Affinity Propagation and other Data Clustering Techniques.

Roy, A. D., 1952. Safety First and the Holding of Assets. *The Econometric Society*, 20(3), pp. 431-449.

Sarker, I.H., 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* 2, 160.

Sass, J., Thös, A.K., 2021. Risk reduction and portfolio optimization using clustering methods, *Econometrics and Statistics*.

Sathya, R., Abraham, A., 2013. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence (IJARAI)*, 2(2).

Schellinger, B., 2020. Optimization of special cryptocurrency portfolios. *The Journal of Risk Finance*, 21(2), pp. 127-157.

Soleymani, F., Vasighi, M., 2022. Efficient portfolio construction by means of CVaR and K-means++ clustering analysis: Evidence from the NYSE. *Int J Fin Econ*, 27, pp. 3679–3693.

Soyster, A. L., 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, 21(5), pp. 1154-1157.

Terraza, V., Toque, C., 2021. Cluster Analysis for Investment Funds Portfolio Optimisation: A Symbolic Data Approach. *Financial Risk Management and Modeling. Risk, Systems and Decisions*. Springer, Cham.

Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Yamai, Y., Yoshida, T., 2005. Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, 29(2005), pp. 997-1015.

Abstract

Investment in financial markets has always been a major concern for investors, particularly in terms of selecting a portfolio composed of various assets available in the market. This issue is especially prominent in emerging markets like the cryptocurrency market, which is exposed to a variety of systematic and unsystematic risks. The selection of an optimal investment portfolio can reduce investment risk while ensuring a suitable return. This enables traditional investors, who have recently taken an interest in the cryptocurrency market and have limited knowledge about it, to make more risk-averse investments in this market.

The current research aims to assist these investors by proposing a multi-stage model. Initially, it clusters the symbols available in the cryptocurrency market using models such as Affinity Propagation, K-Means, and K-Medoids. Then, it selects one or more symbols in each cluster. Finally, it optimizes the investment portfolio using the remaining symbols and methods such as Equal Weighted, Mean-Variance, Mean-CVaR, HRP, and Sparse portfolio.

This research ultimately uses comparative charts to contrast different methods, enabling the investor to choose the best method based on their needs. The results indicate that the model with the best performance in terms of the Sharpe ratio uses the K-Means method for clustering, selects symbols in each cluster based on the highest Sharpe ratio, and uses the HRP method for portfolio selection.

Keywords:

Portfolio Selection, Portfolio Optimization, Cryptocurrency Market, Digital Currencies, Machine Learning, Clustering



K. N. Toosi University Of Technology
Department of Industrial Engineering

Portfolio Optimization in the Cryptocurrency Market Using Machine Learning and Clustering

Alireza Nezhadshamsi

Supervisor

Dr. Amir Abbas Najafi

Advisor

Dr. Majid Mirzaei

Master of Science Thesis in

Industrial Engineering Majoring in Financial Engineering

July 2024