



## انتخاب آنلاین سبد سرمایه‌گذاری به روش تطابق با الگوی طیفی

متین عبدی<sup>۱</sup>

امیرعباس نجفی<sup>۲</sup>

تاریخ پذیرش: ۹۶/۰۹/۱۲

تاریخ دریافت: ۹۶/۰۴/۲۸

### چکیده

با توجه به افزایش حجم و سرعت معاملات در بازارهای مالی امروز، افزایش سرعت در تحلیل‌ها و تصمیم‌گیری‌ها اجتناب ناپذیر است. انجام تحلیل‌های سریع و عاری از خطاهای رفتاری توسط انسان غیر ممکن است. از این رو بازارهای مالی به سمت داد و ستدهای الگوریتمی<sup>۱</sup> در حرکت هستند که در آن‌ها از تکنیک‌هایی از قبیل یادگیری ماشینی و داده‌کاوی<sup>۲</sup> استفاده می‌شود. انتخاب آنلاین سبد سرمایه‌گذاری یکی از تکنیک‌های نوین در داد و ستد الگوریتمی است به این صورت که سرمایه را به تعداد مشخصی سهم تخصیص داده و در ابتدای هر دوره با استفاده از تکنیک‌هایی پرتفو را به روز رسانی می‌کند. در واقع در این تکنیک‌ها انسان در انتخاب پرتفو دخالتی نداشته و الگوریتم نحوه‌ی سرمایه‌گذاری را در هر دوره مشخص می‌کند. در این مقاله الگوریتمی توسعه داده شده است که از اصل تطابق با الگو<sup>۳</sup> در انتخاب آنلاین سبد سرمایه‌گذاری پیروی می‌کند. در این اصل، پرتفو بر اساس الگوهای تاریخی مشابه انتخاب می‌شود که در این مقاله برای یافتن الگوهای تاریخی مشابه از روش خوشه‌بندی طیفی<sup>۴</sup> در داده‌کاوی استفاده شده است. در این خصوص یک مثال عددی با استفاده از ۲۰ سهم فعال تر در بورس نیویورک ارائه شده و نتایج آن با الگوریتم‌های دیگر در این حوزه مقایسه شده است.

**واژه‌های کلیدی:** داد و ستد الگوریتمی، انتخاب آنلاین سبد سرمایه‌گذاری، اصل تطابق با الگو، داده‌کاوی، خوشه‌بندی طیفی.

۱- دانشجوی کارشناسی ارشد مهندسی صنایع-سیستم‌های مالی دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

۲- دانشیار دانشکده مهندسی صنایع دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران (نویسنده مسئول)

aanajafi@kntu.ac.ir

**۱- مقدمه**

امروزه در بازارهای مالی حجم و سرعت معاملات افزایش چشم‌گیری یافته است که نیاز به تکنیک‌های داد و ستد الگوریتمی را افزایش داده است و بدون استفاده از این تکنیک‌ها نمی‌توان همگام با تغییرات بازار پیش رفت. تاکنون بیشتر، روش‌های دستی<sup>۵</sup> برای تحلیل فرصت‌های سرمایه‌گذاری مورد استفاده بوده است که عمده‌ترین این روش‌ها تحلیل‌های تکنیکی و بنیادی می‌باشند. تحلیل‌گران بنیادی بر اساس صورت‌ها و نسبت‌های مالی ارزش ذاتی سهم را محاسبه کرده و آن را با قیمت سهم در بازار مقایسه کرده و بر این اساس در خصوص خرید یا فروش سهم تصمیم‌گیری می‌نمایند. تحلیل‌گران تکنیکی، بر این باورند که عملکرد گذشته هر سهم معیار کافی برای پیش‌بینی عملکرد آینده آن است که در این راستا از نمودارها و اندیکاتورها برای تشخیص الگوها کمک می‌گیرند تا بتوانند قیمت‌های آینده را پیش‌بینی کنند. سرعت این رویکردها به جهت دستی بودن بسیار پایین بوده و نمی‌توانند نیازهای سرمایه‌گذاران در بازارهای مالی امروز را برآورده کنند. برای رفع این نقیصه تکنیک‌های داد و ستد الگوریتمی ارائه شد که یکی از مهمترین آن‌ها انتخاب آنلاین سبد سرمایه‌گذاری است.

انتخاب آنلاین سبد سرمایه‌گذاری به تخصیص سرمایه بین چند سهم و به روز رسانی سبد سهام در هر دوره جهت به حداکثر رساندن بازده سرمایه‌گذاری در مدت طولانی می‌پردازد. در واقع در این روش، الگوریتم، الگویی برای به روز رسانی پرتفو در هر دوره ارائه می‌دهد که تعیین می‌کند در هر دوره چه میزان از سرمایه به کدام سهم تخصیص یابد و بر این اساس در ابتدای هر دوره تعدادی از هر سهم خریداری یا فروخته می‌شود. در انتخاب آنلاین سبد سرمایه‌گذاری، الگوریتم‌هایی ارائه شده است که با ترکیب تئوری‌های اقتصادی و روش‌های داده کاوی و یادگیری ماشینی مختلف به حداکثر کردن بازده سرمایه‌گذاری طی چند دوره می‌پردازد.

**۲- مبانی نظری و مروری بر پیشینه پژوهش**

در ادبیات مدل‌های انتخاب سبد سرمایه‌گذاری، دو اصل تئوریک عمده وجود دارد:

- تئوری اول را مارکوویتز در سال ۱۹۵۲ ارائه کرد که به موازنه‌ی بین بازده انتظاری (میانگین) و ریسک (واریانس) می‌پردازد و برای تصمیم‌گیری تک دوره‌ای مناسب است.
- تئوری دوم را کیلی در سال ۱۹۵۶ ارائه داد که به تئوری رشد سرمایه معروف است و هدف آن حداکثر کردن امید ریاضی لگاریتم بازده سرمایه‌گذاری است و برای تصمیم‌گیری چند دوره‌ای مناسب است. از آنجایی که در انتخاب آنلاین سبد سرمایه‌گذاری تصمیم‌گیری

برای چند دوره‌ی متوالی انجام می‌گیرد، غالباً از تئوری رشد سرمایه کلی در مدل‌سازی‌ها استفاده می‌شود.

الگوریتم‌های ارائه شده در زمینه‌ی انتخاب آنلاین سبد سرمایه‌گذاری از سه اصل اساسی پیروی کرده و بر مبنای این اصول، الگوهایی برای به روز رسانی پرتفو ارائه می‌دهند. اصول یا الگوریتم‌هایی که الگوی به روز رسانی برای پرتفو را ارائه می‌دهند شامل سه دسته اصلی هستند که بر اساس جهت انتقال اوزان دسته‌بندی می‌شوند. یعنی سرمایه از چه استراتژی یا سهمی به کدام استراتژی یا سهم منتقل شود. اولین رویکرد، تبعیت از برنده<sup>۶</sup> است که در هر دوره وزن بیشتری به استراتژی یا سهمی که در گذشته عملکرد بهتری داشته می‌دهد. در مقابل آن، رویکرد دوم، تبعیت از بازنده<sup>۷</sup> است که در هر دوره وزن سرمایه‌گذاری را از استراتژی یا سهمی که در گذشته عملکرد بهتری داشته به استراتژی یا سهمی که عملکرد بدتری داشته منتقل می‌کند. چرا که بر این باور است که سهمی که در گذشته عملکرد بدی داشته در آینده عملکرد خوبی خواهد داشت و این رویکردها بر پایه‌ی اصل بازگشت به میانگین بنا شده‌اند. دسته سوم رویکردهایی بر اساس تطابق با الگو هستند که پرتفو را بر اساس الگوهای تاریخی مشابه طراحی می‌کنند. به طور معمول الگوریتم‌های تطابق با الگو شامل دو گام است : ۱- انتخاب نمونه ۲- بهینه‌سازی پرتفو

جهت انتخاب نمونه، داده‌های اخیر با داده‌های گذشته مقایسه می‌شود و داده‌هایی در گذشته که با توجه به معیارهایی نظیر هیستوگرام، کرنل، نزدیکترین همسایه و همبستگی با داده‌های اخیر شباهت دارند به عنوان نمونه منتخب مشخص می‌شوند. در گام بعد با توجه به نمونه منتخب، پرتفوی بهینه را از طریق توابع هدفی همچون لگاریتم بهینه، نیمه‌لگاریتم بهینه و مارکوویتز به دست می‌آوریم.

همان‌طور که گفته شد اولین اصل ارائه شده، اصل تبعیت از برنده است. کاور (۱۹۹۱) برای اولین بار الگوریتم سبد سرمایه‌گذاری یونیورسال را در این زمینه ارائه کرد. ایده این الگوریتم به این صورت است که در ابتدای دوره بخشی از سرمایه به هر استراتژی تخصیص داده می‌شود و در نهایت عایدی آن‌ها تجمیع می‌یابد. کاور و اوردنتلیچ (۱۹۹۶) همچنین اطلاعات جانبی، نظر خبرگان و اطلاعات بنیادین را وارد مدل کردند. در ادامه کاور و اوردنتلیچ (۱۹۹۸)، فروش استقرایی و خرید اعتباری را نیز به مدل اضافه کردند. همچنین بلوم و کالای (۱۹۹۹)، مدل را با هزینه معاملاتی ارتقا دادند.

دومین اصل ارائه شده در انتخاب آنلاین سبد سرمایه‌گذاری، اصل تبعیت از بازنده است که این اصل بر اساس خاصیت بازگشت به میانگین که اصولاً در داده‌های مالی وجود دارد، بنا نهاده شده

است. بورودین و همکاران (۲۰۰۴) برای اولین بار الگوریتم ناهمبستگی<sup>۸</sup> را در این حوزه ارائه کردند. لی و همکاران (۲۰۱۲) الگوریتم بازگشت به میانگین منفعلانه-فعالانه<sup>۹</sup> را ارائه دادند که از خاصیت بازگشت به میانگین و همچنین یادگیری آنلاین منفعلانه-فعالانه استفاده می‌کند. این روش موازنه‌ای بین بازده پرتفو و ریسک نوسانات ایجاد می‌کند. در ادامه لی و همکاران (۲۰۱۳) الگوریتم بازگشت به میانگین اطمینان وزنی را ارائه دادند که در آن از واریانس اوزان پرتفو نیز استفاده شده است. در این الگوریتم از ترکیب دیدگاه معاملاتی بازگشت به میانگین و روش یادگیری ماشینی اطمینان وزنی استفاده شده است. لی و همکاران (۲۰۱۵) الگوریتم بازگشت به میانگین منفعلانه-فعالانه را به الگوریتم بازگشت به میانگین متحرک ارتقا دادند.

اصل ارائه شده دیگر در حوزه‌ی انتخاب آنلاین سبد سرمایه‌گذاری، اصل تطابق با الگو است. مطالعات انجام شده در این زمینه در دو گام انتخاب نمونه و بهینه سازی پرتفو انجام شده است. گیورفی و شافر (۲۰۰۳) برای اولین بار استراتژی  $B^H$  را معرفی کردند که ترکیبی از دو روش هیستوگرام در گام اول و لگاریتم بهینه در گام دوم است. گیورفی و همکاران (۲۰۰۶)  $B^K$  را معرفی کردند که ترکیبی از دو روش کرنل در گام اول و لگاریتم بهینه در گام دوم است. همچنین گیورفی و همکاران (۲۰۰۸) استراتژی  $B^{NN}$  را ارائه کردند که ترکیبی از دو روش نزدیک‌ترین همسایه در گام اول و لگاریتم بهینه در گام دوم است. لی و همکاران (۲۰۱۱) استراتژی CORN را معرفی کردند. این استراتژی ترکیب روش همبستگی در گام اول و لگاریتم بهینه در گام دوم است و نشان داده شده که از عملکرد تجربی بهتری نسبت به سه الگوریتم قبلی برخوردار است. همچنین گیورفی و همکاران (۲۰۰۷) برای ساده‌سازی محاسبات  $B^K$  استراتژی  $B^S$  را ارائه نمودند که از روش نیمه لگاریتم بهینه در گام دوم بهره می‌جست. اتوساک و وایدا (۲۰۰۷) جهت استفاده از دیدگاه مارکوویتز و موازنه بین میانگین (بازگشت) و واریانس (ریسک)، استراتژی  $B^M$  را ارائه نمودند که ترکیبی از دو روش کرنل در گام اول و مارکوویتز در گام دوم است. گیورفی و وایدا (۲۰۰۸) جهت دخیل نمودن هزینه معاملاتی در محاسبات استراتژی  $B^{GV}$  را معرفی کردند.

## ۲-۱- فرمول بندی مسئله

فرض کنید سرمایه‌گذاری قصد دارد سرمایه خود را در تعدادی سهم ( $m \geq 2$ ) برای افق زمانی  $n$  دوره ( $n \geq 1$ ) سرمایه‌گذاری کند. قیمت پایانی  $m$  سهم در گذشته در هر دوره زمانی مانند  $t$  به وسیله بردار  $p_t = (p_{1,t}, p_{2,t}, \dots, p_{i,t}, \dots, p_{m,t})$  نشان داده می‌شود. همچنین تغییرات قیمتی  $m$  سهم به وسیله بردار  $x_t$  نشان داده می‌شود که نسبت قیمت پایانی هر روز نسبت به روز قبل را نشان می‌دهد و هر مولفه آن به صورت  $x_{t,i} = \frac{p_{t,i}}{p_{t-1,i}}$  به دست می‌آید. بنابراین مبلغ سرمایه‌گذاری شده

در سهم  $i$  در طول دوره  $t$  با نرخ  $x_{t,i}$  افزایش می‌یابد. به این ترتیب  $x_1^n = \{x_1, \dots, x_n\}$  دنباله‌ای از بردارهای تغییرات قیمتی برای  $n$  دوره است که در واقع تشکیل یک ماتریس  $n \times m$  می‌دهد که اندیس  $n$  نشان‌دهنده‌ی دوره زمانی و اندیس  $m$  نشان‌دهنده‌ی سهم است. و همچنین پنجره زمانی بردارهای تغییرات قیمتی بین دوره  $s$  تا  $t$  به صورت  $1 \leq s < e \leq n$   $x_s^e = \{x_s, \dots, x_e\}$  نشان داده می‌شود.

نسبت‌های سرمایه‌گذاری در  $m$  سهم در ابتدای دوره  $t$  به وسیله بردار  $b_t = (b_{t,1}, \dots, b_{t,m})$  نشان داده می‌شود که همه مولفه‌های بردار مثبت بوده و همچنین رابطه  $\sum_{i=1}^m b_{t,i} = 1, \forall t$  برقرار است.

نسبت‌های سرمایه‌گذاری در ابتدای دوره  $t$  به وسیله مشاهده رفتار گذشته بازار یعنی بردار نسبت‌های قیمتی از اولین دوره بررسی شده تا دوره‌ی قبل از  $t$  تعیین می‌شود و به صورت  $b_t = f_t(X_1^{t-1})$  نشان داده می‌شود. در نتیجه  $b_1^n = \{b_1, \dots, b_n\}$  یک استراتژی برای  $n$  دوره و به عنوان خروجی یک استراتژی انتخاب آنلاین سبد سهام است.

با انتخاب پورتفوی  $b_t$  برای دوره  $t$ ، در انتهای این دوره به میزان  $S_t$  بازدهی کسب می‌کنیم که  $S_t = b_t^T x_t = \sum_{i=1}^m b_{t,i} x_{t,i}$  از آنجایی که در ابتدای هر دوره مبلغ به دست آمده را مجدداً سرمایه‌گذاری می‌کنیم، پس از  $n$  دوره و با استفاده از استراتژی  $b_1^n$  پرتفویی با ارزش تجمعی  $S_n$  حاصل می‌شود که ارزش اولیه پرتفو را با نرخ  $\prod_{t=1}^n b_t^T x_t$  افزایش می‌دهد.

$$s_n(b_1^n, x_1^n) = s_0 \prod_{t=1}^n b_t^T x_t = s_0 \prod_{t=1}^n \sum_{i=1}^m b_{t,i} x_{t,i} \quad (1)$$

در رابطه فوق  $S_0$  نشان دهنده ارزش اولیه مبلغ سرمایه‌گذاری شده در پرتفو است. همچنین  $w_n(b_1^n)$  نشان‌دهنده نرخ رشد نمایی سرمایه برای استراتژی  $b_1^n$  است.

$$w_n(b_1^n) = \frac{1}{n} \log s_n(b_1^n) = \frac{1}{n} \sum_{t=1}^n \log b_t^T x_t \quad (2)$$

در این مدل‌ها، تصمیم‌گیرنده باید استراتژی  $b_1^n$  را جهت دستیابی به اهداف مشخص انتخاب کند. در الگوریتم‌های مطرح شده هدف، حداکثرسازی تابع  $S_n$  است. در ابتدای دوره  $t$  تصمیم‌گیرنده بر اساس پنجره زمانی  $X_1^{t-1}$  پورتفوی  $b_t$  را برای دوره بعدی بر اساس استراتژی‌های متفاوت ارائه می‌کند. در انتهای دوره بازدهی پورتفوی انتخاب شده به اندازه  $b_t x_t$  است. این فرآیند تا دوره  $n$  ادامه می‌یابد و در نهایت بازدهی کل استراتژی برابر  $S_n$  خواهد بود.

## ۲-۲- معیارهای ارزیابی مدل

- یک معیار استاندارد برای ارزیابی یک استراتژی انتخاب آنلاین سبد سهام، ارزش تجمعی سبد در انتهای دوره سرمایه‌گذاری است. اگر ارزش اولیه را مساوی ۱ قرار دهیم،  $S_n$  بیانگر بازدهی تجمعی سبد سهام است که هرچه بیشتر باشد برای سرمایه‌گذار مطلوب‌تر است.
- معیار دیگر ارزیابی، درصد عایدی سالیانه<sup>۱۰</sup> است که از رابطه زیر به دست می‌آید:

$$APY = \sqrt[y]{S_n} - 1 \quad (۳)$$

- که  $y$  تعداد سال‌های افق سرمایه‌گذاری است. این معیار میانگین افزایش ارزش توسط این استراتژی در هر سال را اندازه‌گیری می‌کند که طبیعتاً هرچه بیشتر باشد برای سرمایه‌گذار مطلوب‌تر است.
- معیار دیگر که برای ارزیابی ریسک استفاده می‌شود انحراف معیار سالیانه بازدهی‌های سبد سهام است. برای محاسبه انحراف معیار سالیانه، انحراف معیار روزانه بازدهی‌ها را به دست آورده و در  $\sqrt{252}$  ضرب می‌کنیم. (فرض می‌کنیم هر سال ۲۵۲ روز کاری دارد).
  - معیار دیگر نسبت شارپ سالیانه است که باعث ایجاد موازنه‌ای بین ریسک و بازده می‌شود. که این معیار از رابطه زیر محاسبه می‌شود:

$$SR = \frac{APY - R_f}{\sigma_p} \quad (۴)$$

هرچه SR بیشتر باشد، استراتژی دارای ریسک کمتری به ازای بازده مشخص است.

## ۲-۳- اصل تطابق با الگو

همان‌طور که گفته شد، الگوریتم‌های تطابق با الگو شامل دو گام انتخاب نمونه و بهینه‌سازی پرتفو هستند و الگوریتم‌های ارائه شده از ترکیب روش‌های مختلف در این دو گام حاصل شده‌اند. در این بخش تکنیک‌های ارائه شده در هر گام و الگوریتم‌های ارائه شده حاصل از ترکیب این گام‌ها توضیح داده شده‌اند.

### گام اول - انتخاب نمونه:

در این گام، هدف انتخاب نمونه‌های مشابه با دوره مورد تصمیم‌گیری  $t+1$  است. به این منظور در ابتدا باید اندازه‌ی پنجره زمانی  $w$  تعیین شود. سپس از میان مجموعه بردارهای  $X_1^t$  که شامل داده‌های تاریخی تغییرات قیمت است اقدام به پیدا کردن اندیس‌های مشابه می‌کنیم. هدف پیدا

کردن اندیس‌هایی است که در آن اندیس‌ها ( $i$ ) دو پنجره زمانی  $X_{t-w}^{i-1}$  و  $X_{t-w+1}^t$  مشابه باشند. در واقع پنجره زمانی  $X_{t-w+1}^t$  پنجره زمانی اخیر بوده و با یافتن پنجره‌های زمانی مشابه یعنی  $X_{t-w}^{i-1}$  ها می‌توان استنباط کرد که دوره  $t+1$  همانند دوره  $i$  خواهد بود. اندیس‌های مشابه پیدا شده یعنی  $i$ ها را در مجموعه  $C$  قرار داده، سپس به هر یک از اندیس‌ها احتمال  $P_i$  نسبت داده می‌شود که اغلب این احتمال یکنواخت بوده و به صورت  $P_i = \frac{1}{|C|}$  تعریف می‌شود که  $|C|$  تعداد اعضای مجموعه  $C$  است.

برای تشخیص شباهت بین پنجره‌های زمانی چهار روش زیر موجود است:

#### ✓ بر اساس هیستوگرام

در این روش تعدادی پارتیشن<sup>۱۱</sup> به صورت گسسته تعریف می‌شود و در صورتی که پنجره زمانی اخیر ( $X_{t-w+1}^t$ ) و پنجره زمانی تاریخی ( $X_{t-w}^{i-1}$ ) در یک پارتیشن قرار گیرند، آن دو پنجره زمانی مشابه قلمداد می‌شوند.

فضای عددی مورد مطالعه به  $d$  پارتیشن مساوی تقسیم شده است و پارتیشن  $P=A_j$ ,  $j=1,2,\dots,d$  تعریف می‌شود. همچنین تابع  $G(x)=j$  بیانگر شماره پارتیشنی است که  $x$  در آن قرار گرفته است. در صورتی که شرط زیر برقرار باشد دو پنجره زمانی مشابه هستند:

$$C_H(X_1^t, w) = \{w < i < t + 1 : G_t(X_{t-w+1}^t) = G_t(X_{t-w}^{i-1})\} \quad (5)$$

#### ✓ بر اساس کرنل

در این روش فاصله اقلیدسی دو پنجره زمانی محاسبه می‌شود و اگر از مقدار  $\frac{c}{l}$  کمتر باشد، دو پنجره زمانی مشابهند:

$$C_K(X_1^t, w) = \{w < i < t + 1 : \|(X_{t-w+1}^t) - (X_{t-w}^{i-1})\| \leq \frac{c}{l}\} \quad (6)$$

$\frac{c}{l}$  آستانه‌ای برای کنترل تعداد نمونه‌های مشابه است.

#### ✓ بر اساس نزدیک‌ترین همسایه

در این روش  $k$  پنجره زمانی نزدیک‌تر به پنجره‌ی اخیر از نظر فاصله اقلیدسی به عنوان پنجره‌های مشابه قلمداد می‌گردند.

$$C_N(X_1^t, w) = \{w < i < t + 1 : (X_{t-w}^{i-1}) \text{ is among the } k \text{ NNs of } (X_{t-w+1}^t)\} \quad (7)$$

### ۷ بر اساس همبستگی

در این روش تشابه بر اساس ضریب همبستگی پنجره زمانی اخیر و پنجره زمانی تاریخی است.

$$C_C(X_1^t, w) = \left\{ w < i < t + 1 : \frac{\text{cov}(X_{i-w}^{i-1}, X_{t-w+1}^t)}{\text{std}(X_{i-w}^{i-1})\text{std}(X_{t-w+1}^t)} \geq \rho \right\} \quad (8)$$

### گام دوم - بهینه‌سازی پرتفو

گام دوم در الگوریتم‌های تطابق با الگو انتخاب پرتفوی بهینه بر اساس مجموعه  $C$  است. دو رویکرد عمده در انتخاب پرتفو وجود دارد: تئوری رشد سرمایه کلی و تئوری مارکویتز. در ادامه چند تکنیک بر اساس این دو رویکرد عمده برای انتخاب پرتفو معرفی خواهد شد. گیورفی و همکاران روش لگاریتم بهینه را برای انتخاب پرتفو بر اساس تئوری رشد سرمایه کلی معرفی کردند. تابع سود در این روش به صورت زیر است:

$$U_L(b; C(x_1^t)) = E\{\log b \cdot x \mid x_i, i \in C(x_1^t)\} = \sum_{i \in C(x_1^t)} P_i \log b \cdot x_i \quad (9)$$

که  $P_i$  بیانگر احتمال وقوع هر یک از اعضای مجموعه  $C$  است. در صورتی که احتمال یکنواخت باشد رابطه بالا به شکل زیر در می‌آید:

$$U_L(b; C(x_1^t)) = \sum_{i \in C(x_1^t)} \log b \cdot x_i \quad (10)$$

گیورفی و همکاران برای حل مشکل محاسباتی روش بالا، روش نیمه لگاریتم بهینه را ارائه دادند که در تابع سود از تقریب لگاریتم استفاده می‌کند.

$$U_S(b; C(x_1^t)) = E\{f(b \cdot x) \mid x_i, i \in C(x_1^t)\} = \sum_{i \in C(x_1^t)} P_i f(b \cdot x_i) \quad (11)$$

که تابع  $f$  مرتبه دوم بسط تیلور  $\log z$  است. در صورتی که  $z=1$  باشد خواهیم داشت:

$$f(z) = z - 1 - \frac{1}{2}(z - 1)^2 \quad (12)$$

در صورتی که تابع احتمال یکنواخت باشد تابع سود به شکل زیر خواهد بود:

$$U_S(b; C(x_1^t)) = \sum_{i \in C(x_1^t)} f(b \cdot x_i) \quad (13)$$



اتوساک و وایدا روشی بر مبنای رویکرد مارکویتز ارائه دادند. ایده اصلی این رویکرد موازنه بین میانگین (بازدهی) و واریانس (ریسک) است. تابع سود مارکویتز به صورت زیر تعریف می‌شود:

$$U_M(b; C(x_1^t)) = E\{b \cdot x | x_i, i \in C(x_1^t)\} - \lambda \text{Var}\{b \cdot x | x_i, i \in C(x_1^t)\} - \lambda E\{b \cdot x | x_i, i \in C(x_1^t)\}^2 \quad (14)$$

که  $\lambda$  پارامتر موازنه است.

برای حل مسائل با هزینه معاملاتی، گیورفی و وایدا<sup>۱۳</sup>، تابع سود زیر<sup>۱۳</sup> را ارائه نمودند که دربرگیرنده هزینه معاملاتی است:

$$U_T(b; C(x_1^t)) = E\{\log(b \cdot x) + \log c(b_t \cdot b \cdot x_t)\} \quad (15)$$

که  $C(\cdot) \in (0,1)$  فاکتور هزینه معاملاتی است و بیانگر نسبت باقی‌مانده پس از کسر هزینه معاملاتی است. اگر تابع احتمال به صورت یکنواخت باشد تابع سود به این صورت خواهد بود:

$$U_T(b; C(x_1^t)) = \sum_{i \in C(x_1^t)} (\log b \cdot x + \log c(b_t \cdot b \cdot x_t)) \quad (16)$$

اگر مجموعه  $C$  غیرتهی باشد از هر یک از روش‌های بالا بسته به محدودیت‌ها و یا رویکرد برای انتخاب پرتفوی بهینه می‌توان استفاده کرد ولی در صورتی که مجموعه  $C$  تهی باشد پرتفوی بهینه می‌تواند یکنواخت و یا پرتفوی قبلی باشد.

### گام سوم - ترکیب گام‌های قبل

الگوریتم‌های ارائه شده در زمینه تطابق با الگو از ترکیب روش‌های توضیح داده شده حاصل شده‌اند. در جدول (۱) ترکیب روش‌های انتخاب نمونه و بهینه‌سازی پرتفو که تا کنون انجام گرفته نشان داده شده است. خانه‌های (-) بیانگر این است که تاکنون الگوریتمی در آن زمینه مطرح نشده است.

الگوریتم CORN شامل دو الگوریتم CORN-U و CORN-K می‌باشد. در الگوریتم CORN-U تابع احتمال  $P_1$  که به اندیس‌های مشابه نسبت داده می‌شود تابع یکنواخت بوده و در واقع به هر یک از اندیس‌ها وزن یکسان داده می‌شود. اما در الگوریتم CORN-K، نمونه که شباهت بیشتری به نمونه اخیر دارند انتخاب شده و به این  $k$  نمونه وزن‌های مساوی نسبت داده می‌شود.

جدول ۱- خلاصه‌ی تحقیقات صورت گرفته در رویکردهای تطابق با الگو

تکنیک‌های انتخاب نمونه				بهینه‌سازی پورتفو
همبستگی	نزدیک‌ترین همسایه	کرنل	هیستوگرام	
CORN: C <sub>C</sub> +U <sub>L</sub>	B <sup>NN</sup> : C <sub>N</sub> +U <sub>L</sub>	B <sup>K</sup> : C <sub>K</sub> +U <sub>L</sub>	B <sup>H</sup> : C <sub>H</sub> +U <sub>L</sub>	لگاریتم بهینه
-	-	B <sup>S</sup> : C <sub>K</sub> +U <sub>S</sub>	-	نیمه لگاریتم بهینه
-	-	B <sup>M</sup> : C <sub>K</sub> +U <sub>M</sub>	-	مارکویتز
-	-	B <sup>GV</sup> : C <sub>K</sub> +U <sub>R</sub>	-	GV

### ۳- روش شناسی پژوهش

تفاوت عمده‌ای که پژوهش پیش رو با مطالعات گذشته دارد این است که در گام انتخاب نمونه از خوشه بندی طیفی استفاده شده است.

خوشه بندی یکی از روش‌های داده کاوی است که به منظور افراز داده‌ها بر اساس تشابه آن‌ها مورد استفاده قرار می‌گیرد. یک خوشه به مجموعه‌ای از داده‌ها اطلاق می‌شود که تا حد امکان دارای بیشترین تشابه نسبت به هم هستند. خوشه بندی در واقع افراز داده‌ها به گروه‌هایی است که اعضای هر گروه از جهات خاصی به یکدیگر شباهت داشته و با اعضای سایر خوشه‌ها کمترین شباهت را دارند. معیار شباهت در اینجا همبستگی است. با محاسبه‌ی همبستگی بین دو داده می‌توان فهمید تا چه اندازه این دو داده به هم شباهت دارند و بر این اساس می‌توان آن‌ها را در یک خوشه قرار داد.

یکی از تکنیک‌های خوشه بندی، خوشه بندی طیفی است که از مقادیر و بردارهای ویژه‌ی ماتریس تشابه داده‌ها برای خوشه بندی استفاده می‌کند. ورودی این الگوریتم  $n$  داده و مقدار  $k$  که تعداد خوشه‌های خروجی را مشخص می‌کند، می‌باشد. ماتریس تشابه داده‌ها با استفاده از فواصل بین داده‌ها تشکیل شده و شامل ارزیابی کمی از تشابه نسبی هر جفت از نقاط مجموعه است. در ابتدا ماتریس  $A$  با استفاده از ماتریس تشابه داده‌ها ساخته شده و  $k$  بردار ویژه‌ای که متناظر با بیشترین مقادیر ویژه هستند، به دست می‌آید. سپس بردارهای ویژه به صورت ستونی در کنار هم قرار گرفته و تشکیل ماتریس می‌دهند. پس از نرمال‌سازی ماتریس به دست آمده هر سطر از ماتریس جدید نماینده یک نقطه در فضای  $k$  بعدی خواهد بود. این نقاط جدید به دست آمده را با کمک یکی از روش‌های استاندارد خوشه بندی از جمله  $k$ -Means خوشه بندی می‌کنیم. هر نقطه جدید به دست آمده نماینده یکی از داده‌های اصلی است و با توجه به این که نقطه جدید در کدام خوشه قرار گرفته است، داده اصلی متناظر با آن هم در همان خوشه قرار خواهد گرفت. در روش  $k$ -Means تعداد  $k$  نمونه به طور تصادفی از میان کل نمونه‌ها به عنوان نماینده‌ی  $k$  خوشه انتخاب

خواهند شد. با استفاده از معیار تشابه، شباهت هر یک از نمونه‌های باقی‌مانده را با  $k$  نماینده محاسبه می‌کنیم و نمونه مورد نظر به هر یک شبیه‌تر بود، به عضویت آن خوشه در می‌آید. پس از آن برای هر خوشه، با محاسبه‌ی میانگین میان اعضای خوشه نماینده‌ی جدیدی انتخاب می‌گردد. این فرآیند می‌تواند تا هنگامی که دیگر هیچ یک از نمونه‌ها خوشه‌های خود را تغییر ندهند، ادامه پیدا کند.

در این پژوهش از اصل تطابق با الگو استفاده شده است. بردار نسبت قیمتی برای تعدادی سهم و به ازای تعداد مشخصی دوره زمانی در یک ماتریس قرار گرفته که بخشی از آن به عنوان داده‌های تاریخی<sup>۱۴</sup> و بخش دیگری برای تست مدل استفاده می‌شود. همان طور که بیان شد این اصل شامل دو گام است: انتخاب نمونه و بهینه سازی پرتفو.

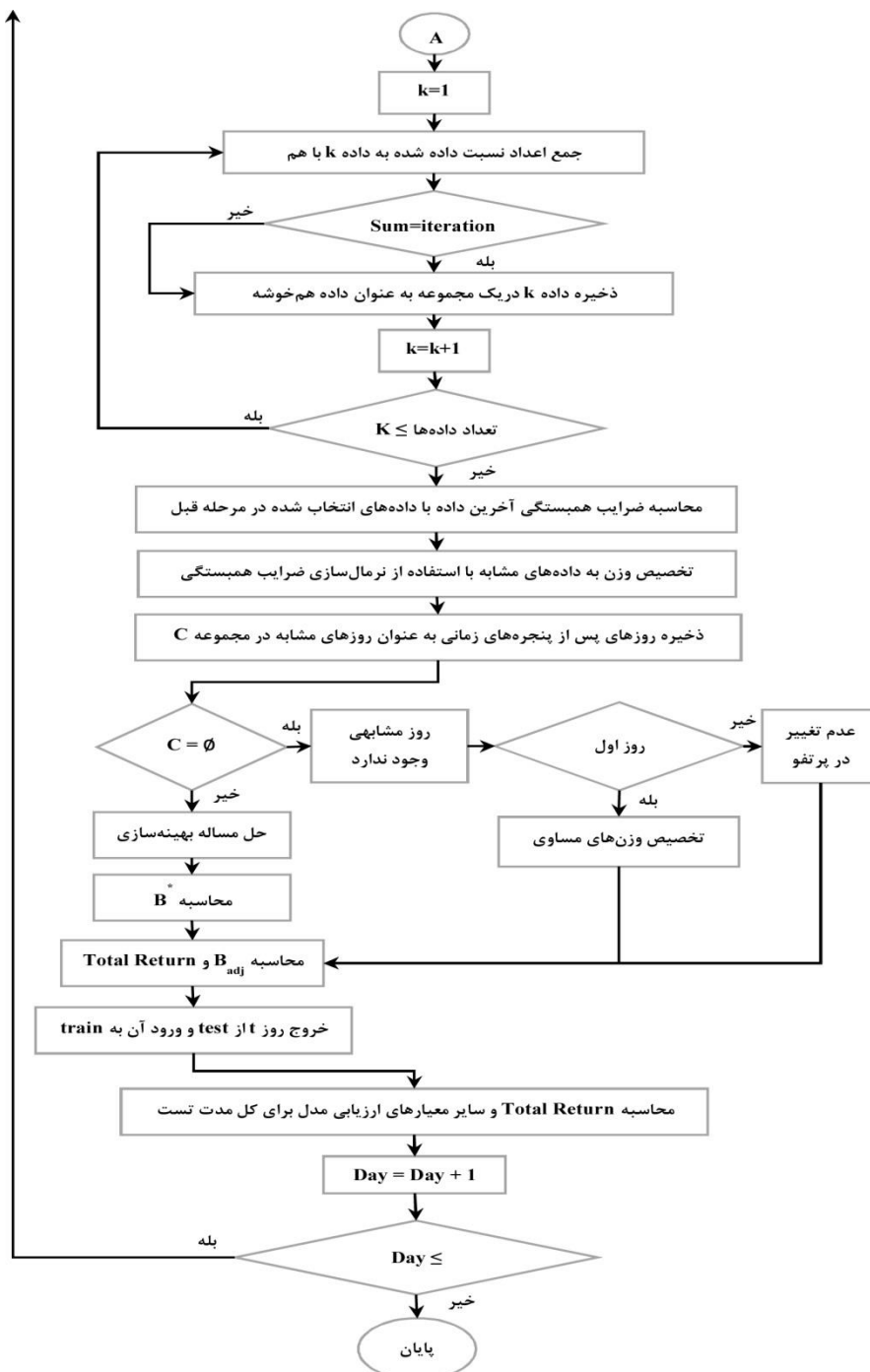
در این پژوهش برای انتخاب نمونه در گام اول از روش خوشه بندی طیفی استفاده شده است. لی و همکاران ثابت کردند که بهترین معیار تشابه در انتخاب نمونه همبستگی است و بر همین اساس خوشه بندی انجام شده در این پژوهش بر اساس همبستگی است. ابتدا اندازه پنجره زمانی توسط کاربر تعیین می‌شود و الگوریتم، پنجره زمانی اخیر و سایر پنجره‌های زمانی را بر اساس این اندازه تولید می‌کند. سپس تعداد خوشه‌ی بهینه برای انجام فرآیند خوشه بندی تعیین می‌شود. به این صورت که خوشه بندی با تعداد خوشه‌های ۲ الی جذر تعداد داده‌ها، به ازای هر کدام به تعداد زیاد تکرار می‌شود و تعداد خوشه‌ای که تابع سیلهوت<sup>۱۵</sup> را حداکثر کند تعداد خوشه‌ی بهینه خواهد بود.

در مرحله‌ی بعد، الگوریتم خوشه‌بندی را با تعداد خوشه‌ی بهینه‌ای که در مرحله‌ی قبل تعیین شد به تعداد زیاد انجام می‌دهد و داده‌هایی که در  $\alpha\%$  مواقع با آخرین داده، یعنی پنجره زمانی اخیر، هم‌خوشه بوده اند به عنوان داده‌های مشابه و عضو مجموعه‌ی  $C$  انتخاب می‌شوند. پس از انتخاب نمونه‌های مشابه، برای تعیین  $P_i$ ها از همبستگی میان داده‌های مشابه و داده‌ی اخیر استفاده می‌شود به این صورت که هر داده‌ای که همبستگی بیشتری با داده‌ی اخیر داشته باشد وزن بیشتری خواهد داشت و البته لازم است که جمع اوزان برابر با یک باشد. به این ترتیب با نرمال سازی همبستگی داده‌های مشابه با داده‌ی اخیر وزن هر داده مشخص می‌شود. در گام دوم پرتفوی بهینه را بر اساس نمونه‌های مشابه انتخاب می‌کنیم. در اینجا از روش لگاریتم بهینه برای انتخاب پرتفو استفاده شده است و تابع هدف به صورت زیر است:

$$U_L(b; C(x_1^t)) = E\{\log b \cdot x \mid x_i, i \in C(x_1^t)\} = \sum_{i \in C(x_1^t)} P_i \log b \cdot x_i \quad (17)$$

اگر در گام اول، یعنی انتخاب نمونه‌های مشابه، هیچ داده‌ی مشابهی با داده‌ی اخیر پیدا نشود، الگوریتم اعلام می‌کند که روز مشابهی یافت نشده و اگر در روز اول استفاده از الگوریتم باشیم، نسبت‌های برابر به سهم‌ها تعلق می‌گیرد، در غیر این صورت پرتفوی دوره‌ی قبل تغییری نمی‌کند و همان پرتفو حفظ می‌شود.

در ابتدای هر دوره، نسبت‌های به دست آمده از بخش بهینه‌سازی ( $B^*$ ) برای سرمایه‌گذاری چاپ شده و در انتهای دوره با توجه به تغییرات قیمت رخ داده شده در سهم، نسبتی از سرمایه که در هر سهم قرار دارد  $B_{adj}$  محاسبه و چاپ می‌شود. در واقع زمانی که نمونه‌ی مشابهی در گذشته یافت نشده و پرتفو بدون تغییر باقی می‌ماند،  $B_{adj}$  دوره قبل به عنوان پرتفوی دوره‌ی آتی اعلام می‌شود. همچنین در انتهای هر دوره، Total Return تا آن دوره محاسبه و چاپ شده و دوره‌ی مذکور از بخش تست مدل خارج شده و به داده‌های تاریخی اضافه می‌شود. در نهایت پس از پایان الگوریتم برای تمام مدت تست، معیارهای ارزیابی مذکور محاسبه می‌شود. در شکل زیر خلاصه‌ای از الگوریتم به صورت نموداری برای درک بهتر ارائه شده است:



#### ۴- یافته‌های پژوهش

در این بخش یک مثال عددی برای انتخاب آنلاین سبد سرمایه‌گذاری با استفاده از الگوریتم توسعه داده شده ارائه می‌شود. بدین منظور ۲۰ سهم فعال‌تر از بورس نیویورک انتخاب شده و قیمت‌های پایانی یک سال اخیر آن‌ها، یعنی از آوریل ۲۰۱۶ تا آوریل ۲۰۱۷، استخراج شده است. برای اجرای مدل، نیمی از داده‌ها در بخش تست مدل و نیم دیگر در بخش داده‌های تاریخی قرار گرفته‌اند. به عبارتی، با استفاده از الگوریتم برای ۶ ماه نسبت‌های سرمایه‌گذاری در هر روز تعیین شده و بازدهی پرتفو در ۶ ماه به دست آمده است. در این مثال دوره‌ها روزانه بوده و اندازه‌ی پنجره زمانی ۴ روز در نظر گرفته شده است. علت انتخاب سهم‌های فعال‌تر در نظر نگرفتن نقدشوندگی سهم در مدل است و برای کاهش اثر آن از سهم‌های فعال‌تر که نقدشوندگی بالاتری دارند استفاده شده است.

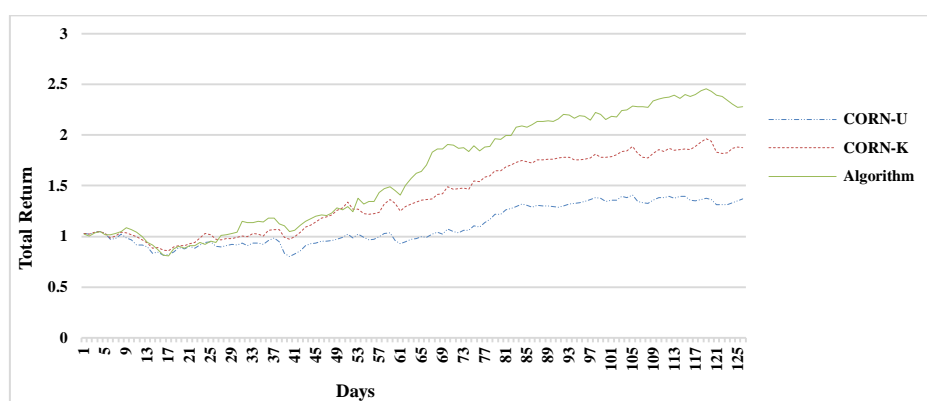
با توجه به توضیحات ارائه شده، نتایج حاصل از الگوریتم در جدول (۲) ارائه شده است. همچنین با توجه به اینکه الگوریتم ارائه شده الهام گرفته از الگوریتم CORN می‌باشد، نتایج حاصل از الگوریتم با الگوریتم‌های CORN-U و CORN-K مقایسه شده تا بهبود آن نسبت به هر دو الگوریتم قابل مشاهده باشد.

جدول ۲- مقایسه نتایج حاصل از مدل پیشنهادی و الگوریتم‌های موجود

CORN-K	CORN-U	روش پیشنهادی	معیارهای ارزیابی
۱,۸۸	۱,۳۷	۲,۲۸	ارزش تجمعی سبد (Sn)
۲,۵۴	۰,۸۸	۴,۱۹	درصد عایدی سالیانه (APY)
۰,۳۹	۰,۴۴	۰,۴۶	انحراف معیار سالیانه بازدهی‌های سبد سهام
۶,۴۲	۱,۹۲	۹,۰۳	نسبت شارپ سالیانه

با توجه به جدول، ارزش تجمعی سبد در طی مدت شش ماه تست با استفاده از روش پیشنهادی برابر با ۲,۲۸ است که همان طور که از جدول مشخص است به میزان قابل توجهی با الگوریتم‌های ارائه شده در گذشته تفاوت دارد و این حاکی از بهبود الگوریتم ارائه شده نسبت به الگوریتم‌های پیشین است. به تبع ارزش تجمعی سبد، درصد عایدی سالیانه الگوریتم ارائه شده نیز به میزان قابل توجهی بیشتر از الگوریتم‌های گذشته و برابر با ۴,۱۹ است. در ارزیابی مطلوبیت یک الگوریتم، علاوه بر بازدهی، به ریسک الگوریتم نیز باید توجه داشت که در اینجا از انحراف معیار سالیانه بازدهی‌های سبد سهام به عنوان معیار اندازه‌گیری ریسک استفاده شده است. همان طور که

در جدول مشخص است، ریسک الگوریتم ارائه شده کمی بیشتر از ریسک الگوریتم‌های پیشین است اما مهمترین معیار برای اندازه گیری مطلوبیت یک الگوریتم، بازده تعدیل شده نسبت به ریسک است که در اینجا از نسبت شارپ سالیانه به این منظور استفاده شده است. نسبت شارپ سالیانه الگوریتم ارائه شده برابر با ۹,۰۳ است که از الگوریتم‌های ارائه شده در گذشته بالاتر است و این به معنی افزایش مطلوبیت الگوریتم ارائه شده در این مقاله نسبت به الگوریتم‌ها پیشین است.



نمودار ۱- مقایسه نتایج بازدهی تجمعی حاصل از مدل پیشنهادی و الگوریتم‌های موجود

با توجه به نمودار(۱)، روند بازدهی تجمعی حاصل از الگوریتم صعودی بوده و همان طور که از شکل پیداست از الگوریتم‌های پیشین بیشتر است.

##### ۵- نتیجه‌گیری و بحث

با توجه به نیاز روز افزون به استفاده از روش‌های دادوستد الگوریتمی به جای استفاده از روش‌های سنتی و تحلیل‌های تکنیکی و بنیادی، ارائه‌ی الگوریتم‌هایی در این زمینه بسیار مفید و کارآمد است. در این پژوهش الگوریتمی در حوزه انتخاب آنلاین سبد سرمایه‌گذاری که از زمینه‌های مهم در دادوستد الگوریتمی است، توسعه داده شده است. این الگوریتم از اصل تطابق با الگو پیروی می‌کند و در این پژوهش سعی شده است کاستی‌های الگوریتم‌های پیشین مرتفع گردد. الگوریتم ارائه شده در این پژوهش الهام گرفته از الگوریتم CORN است که تفاوت عمده روش پیشنهادی با الگوریتم CORN در بخش انتخاب نمونه است. در الگوریتم CORN نمونه مشابه به وسیله یک الگوریتم جست و جو ساده بر اساس همبستگی انتخاب می‌شود به این ترتیب که همبستگی هر

یک از نمونه‌ها که با پنجره زمانی اخیر بیش از حد آستانه‌ای که توسط کاربر تعیین می‌شود باشد، آن نمونه به عنوان نمونه مشابه انتخاب می‌شود. در الگوریتم ارائه شده در این پژوهش هدف انتخاب نمونه‌ای مناسب‌تر جهت بهینه سازی پرتفو است. بدین منظور از روش خوشه بندی طیفی که یکی از تکنیک‌های نوین در داده کاوی است، در بخش انتخاب نمونه استفاده شده است تا نمونه‌ای مناسب تر انتخاب شود. در این روش نیازی به تعیین حد آستانه توسط کاربر نیست و این باعث کاهش خطای انسانی در الگوریتم می‌شود. با توجه به نتایج حاصل از الگوریتم، انتخاب نمونه مشابه با استفاده از روش خوشه بندی منجر به بهبود نتایج الگوریتم شده است و این بدین معناست که نمونه مناسب تری با استفاده از این روش انتخاب شده است. البته هنوز ایراداتی به مدل وارد است که با مطالعات بیشتر می‌توان آن‌ها را رفع کرد. برای مثال در نظر نگرفتن هزینه معاملاتی و نقدشوندگی از مهم‌ترین ایرادات وارد بر این الگوریتم است.



## فهرست منابع

- \* Blum, A., & Kalai, A. (1999). Universal portfolios with and without transaction costs. *Machine Learning*, 35(3), 193-205.
- \* Borodin, A., El-Yaniv, R., & Gogan, V. (2004). Can we learn to beat the best stock. *Journal of Artificial Intelligence Research*, 21, 579-594.
- \* Cover, T. M. (1991). Universal portfolios. *Mathematical finance*, 1(1), 1-29.
- \* Cover, T. M., & Ordentlich, E. (1996). Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(2), 348-363.
- \* Cover, T., & Ordentlich, E. (1998). Universal portfolios with short sales and margin. In *IEEE international symposium on information theory*.
- \* Györfi, L., & Schäfer, D. (2003). Nonparametric prediction. *Advances in learning theory: methods, models and applications*, 339, 354.
- \* Györfi, L., Lugosi, G., & Udina, F. (2006). NONPARAMETRIC KERNEL-BASED SEQUENTIAL INVESTMENT STRATEGIES. *Mathematical Finance*, 16(2), 337-357.
- \* Györfi, L., Urbán, A., & Vajda, I. (2007). Kernel-based semi-log-optimal empirical portfolio selection strategies. *International Journal of Theoretical and Applied Finance*, 10(03), 505-516.
- \* Györfi, L., Udina, F., & Walk, H. (2008). Nonparametric nearest neighbor based empirical portfolio selection strategies. *Statistics & Decisions International mathematical journal for stochastic methods and models*, 26(2), 145-157.
- \* Györfi, L., & Vajda, I. (2008, October). Growth optimal investment with transaction costs. In *International Conference on Algorithmic Learning Theory* (pp. 108-122). Springer Berlin Heidelberg.
- \* Kelly, J. (1956). A new interpretation of information rate. *IRE Transactions on Information Theory*, 2(3), 185-189.
- \* Li, B., Hoi, S. C., & Gopalkrishnan, V. (2011). Corn: Correlation-driven nonparametric learning approach for portfolio selection. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 21.
- \* Li, B., Zhao, P., Hoi, S. C., & Gopalkrishnan, V. (2012). PAMR: Passive aggressive mean reversion strategy for portfolio selection. *Machine learning*, 87(2), 221-258.
- \* Li, B., Hoi, S. C., Zhao, P., & Gopalkrishnan, V. (2013). Confidence weighted mean reversion strategy for online portfolio selection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(1), 4.
- \* Li, B., Hoi, S. C., Sahoo, D., & Liu, Z. Y. (2015). Moving average reversion strategy for on-line portfolio selection. *Artificial Intelligence*, 222, 104-123.
- \* Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1), 77-91.
- \* Ottucsák, G., & Vajda, I. (2007). An asymptotic analysis of the mean-variance portfolio selection. *Statistics & Decisions*, 25(1/2007), 63-86

## یادداشت‌ها

---

- <sup>1</sup> Algorithmic Trading
- <sup>2</sup> Data Mining
- <sup>3</sup> Pattern Matching
- <sup>4</sup> Spectral Clustering
- <sup>5</sup> Manual Analysis
- <sup>6</sup> Follow The Winner
- <sup>7</sup> Follow The Loser
- <sup>8</sup> Anticorrelation
- <sup>9</sup> Passive-Aggressive Mean Reversion (PAMR)
- <sup>10</sup> Annualized Percentage Yield
- <sup>11</sup> Partitions
- <sup>12</sup> Taylor Expansion
- <sup>13</sup> GV-type
- <sup>14</sup> Train
- <sup>15</sup> Silhouette