

کاربرد تئوری مجموعه‌های راف برای پیش‌بینی قیمت سهام (مطالعه موردی):

بانک صادرات ایران^۱

علیرضا سارنج^۲، تورج کریمی^۳، مجید شهرامی بابکان^۴

چکیده

در این پژوهش روشی مبتنی بر تئوری مجموعه‌های راف و با استفاده از شاخص‌های تحلیل تکنیکی جهت پیش‌بینی قیمت سهام ارائه شده است. تئوری مجموعه‌های راف دارای مزایای متعددی است که مهم‌ترین مزیت آن در تحلیل داده‌ها این است که به هیچ‌گونه اطلاعات اضافی اولیه در مورد داده‌ها نیاز ندارد. در مدل پیشنهادی، تعدادی از شاخص‌های تکنیکال برای داده‌های مربوط به بانک صادرات ایران در طول یک سال محاسبه و به‌عنوان مشخصه‌های شرطی در جدول تصمیم مورد استفاده قرار گرفته و نوسان قیمت سهام در روز بعد نیز به‌عنوان مشخصه تصمیم انتخاب می‌شود. لازم به ذکر است که با استفاده از آنالیز ماتریس همبستگی، شاخص‌های با بیشترین همبستگی با مشخصه تصمیم انتخاب می‌گردند. سپس با استفاده از تئوری مجموعه‌های راف و ترکیب روش‌های مختلف گسسته سازی داده‌ها و تولید بی زائده بر اساس داده‌های یادگیری، قواعد پیش‌بینی استخراج و قدرت پیش‌بینی روش‌های مختلف بر اساس داده‌های کنترل محاسبه شد. در این مطالعه داده‌های شش سال متوالی (۱۳۸۷/۰۵/۰۵ لغایت ۱۳۹۴/۰۴/۲۴) بانک صادرات مورد استفاده قرار گرفته است. بررسی قدرت پیش‌بینی این روش و مقایسه بازده حاصل از استفاده از آن و روش خرید و نگهداری، مزیت استفاده از مجموعه‌های راف را آشکار می‌نماید. مقایسه نتایج حاصل از اعمال روش‌ها بر روی داده‌های مربوطه نشان می‌دهد که بازده حاصل از استراتژی خرید و نگهداری ۳۳ ریال و بازده حاصل از مدل ۱۸۲ ریال به ازای هر سهم می‌باشد. همچنین استفاده از داده‌های سال‌های مختلف با روند قیمتی متفاوت به‌عنوان ورودی مدل و دستیابی به نتایج رضایت‌بخش، می‌تواند دلیلی امیدوارکننده برای استفاده از این روش و توسعه آن در پیش‌بینی قیمت سهام باشد.

واژه‌های کلیدی: پیش‌بینی قیمت سهام، تئوری مجموعه‌های راف، مشخصه‌های شرطی و مشخصه‌های

تصمیم، گسسته سازی داده‌ها، استخراج قوانین

طبقه‌بندی موضوعی: .. ، .. ، .. ، ..

۱. کد ... مقاله:

۲. استادیار دانشکده مدیریت و حسابداری پردیس فارابی، دانشگاه تهران، قم، ایران، نویسنده مسئول،

۳. استادیار دانشکده مدیریت و حسابداری پردیس فارابی، دانشگاه تهران، قم، ایران،

۴. دانشجوی دکتری، دانشکده مدیریت و حسابداری پردیس فارابی، دانشگاه تهران، قم، ایران،

مقدمه

در بازار سهام، به دلیل عوامل پیچیده مؤثر بر بازار و روابط غیرخطی که در دوره‌های مختلف قیمت سهام وجود دارد، پیش‌بینی روند قیمت سهم خیلی مشکل می‌باشد؛ بنابراین، از زمانی که اولین بازار گشایش یافت، تلاش‌های زیادی در راستای روش‌های پیش‌بینی قیمت سهام انجام گرفته است. در فضای پیش‌بینی قیمت سهام، روش تحلیل تکنیکی یکی از روش‌های تحلیل مقدماتی است که از سوی سرمایه‌گذاران جهت اخذ تصمیمات سرمایه‌گذاری استفاده شده و بسیاری از پژوهشگران به منظور افزایش بازدهی سرمایه‌گذاری بر روی آن تمرکز نموده‌اند. بعلاوه، روش تحلیل تکنیکی توانایی پیش‌بینی سمت و سوی قیمت در آینده را با مطالعه داده‌های گذشته بازار، حجم و قیمت اولیه سهم دارد. در تحلیل تکنیکی فرض بر این است که حجم و قیمت سهم دو عامل تعیین‌کننده در جهت و رفتار یک سهم خاص یا بازار می‌باشد و شاخص‌های تکنیکال که از فرمول‌های ریاضی با استفاده از قیمت و حجم به دست می‌آیند، می‌توانند به منظور پیش‌بینی قیمت بکار برده شوند. همچنین داده‌هایی برای سرمایه‌گذاران تأمین می‌کند که آن‌ها را قادر می‌سازد تا زمان‌بندی خرید و فروش سهم را تعیین نمایند. علاوه بر روش‌های تحلیل تکنیکی، بسیاری از مدل‌های پیش‌بینی عددی سنتی توسط پژوهشگران مالی نظیر مدل‌های سری زمانی خانواده و پیشنهاد شد. در دهه‌های اخیر، پژوهشگران روش‌های دیگری برای پیش‌بینی مالی به کار گرفته‌اند. الگوریتم هوش مصنوعی که در سال ۱۹۹۰ توسط کینوتو و همکاران معرفی شد، شامل یک سیستم پیش‌بینی برای بازار سهام با استفاده از شبکه عصبی بود. نیکوپولوس و فلراث (۱۹۹۴) الگوریتم ژنتیک و شبکه عصبی را به منظور ایجاد یک سیستم خبره برای تصمیم‌گیری‌های سرمایه‌گذاری ترکیب نمودند. کیم و هان (۲۰۰۰) یک روش الگوریتم ژنتیک را به منظور گسسته‌سازی مشخصه‌ها و تعیین وزن‌های ارتباطی برای شبکه‌های عصبی مصنوعی به منظور پیش‌بینی شاخص قیمت سهم پیشنهاد نمودند. هورانگ و یو (۲۰۰۶) یک شبکه عصبی گذشته‌نگر را به کار گرفتند تا روابط فازی را در سری‌های زمانی فازی برای پیش‌بینی قیمت سهام استخراج نمایند. راج (۲۰۰۷) شبکه عصبی و مدل سری زمانی را برای پیش‌بینی نوسانات شاخص قیمت سهام تلفیق نمود. از موارد ذکر شده در بالا، می‌توان سه ایراد اساسی در مدل‌ها و روش‌های پیش‌بینی آن‌ها گرفت: (۱) تحلیل‌گران بازار سهام و مدیران صندوق‌ها، شاخص‌های تکنیکال متعددی را بر اساس تجربه شخصی خودشان به منظور پیش‌بینی روندهای بازار سهام به کار می‌برند که می‌تواند منجر به قضاوت‌های مخدوش از علائم بازار شود؛ (۲) برای بیشتر روش‌های آماری، برخی فرضیات در مورد متغیرهای مورد استفاده در تحلیل وجود دارد که نمی‌توان آن‌ها را در مورد سایر داده‌هایی که از آن توزیع‌های آماری پیروی نمی‌کنند بکار برد؛ و (۳) شبکه عصبی مصنوعی یک روش جعبه سیاه است و قواعدی که از آن استخراج می‌شوند به راحتی قابل درک نمی‌باشند.

بررسی موارد ذکر شده نیاز به استفاده از روش‌های نوین با قابلیت‌های بیشتر را نمایان می‌نماید که در این مطالعه به کاربرد مجموعه‌های راف در پیش‌بینی قیمت سهام پرداخته‌ایم.

مبانی نظری و مروری بر پیشینه پژوهش

تئوری مجموعه راف^۱ توسط پژوهشگر لهستانی به نام پاولاک در سال ۱۹۸۲ معرفی گردید. از آنجا که پژوهش‌های اولیه در این زمینه به زبان لهستانی منتشر شد، لذا نظر پژوهشگران علوم کامپیوتر و ریاضیات در سطح بین‌المللی را جلب نکرد و توجه چندانی به این تئوری صورت نگرفت. در اواخر دهه ۸۰ این تئوری در سطح جهانی مطرح شد و در سال ۱۹۹۲ اولین کنفرانس علمی تئوری مجموعه راف به صورت بین‌المللی در لهستان برگزار شد. در سال ۱۹۹۵ از سوی ...^۲ به‌عنوان موضوع نوظهور در علم کامپیوتر معرفی شد و در سال ۱۹۹۸ مجله علم اطلاعات^۳ مجموعه‌ای از مقالات مربوط به مجموعه‌های راف را چاپ نمود. از آن پس مجموعه‌های راف به‌عنوان یک ابزار محاسباتی جدید برای برخورد با شرایط مبهم و عدم قطعیت شناخته شد که می‌تواند برای تحلیل اطلاعات غیردقیق، متناقض و ناکامل بکار گرفته شود. کاربرد مجموعه‌های راف را می‌توان به تحلیل‌های تصمیم‌گیری و غیر تصمیم‌گیری تقسیم نمود. تحلیل‌های غیر تصمیم‌گیری به‌طور عمده فشرده نمودن اطلاعات، تقلیل اطلاعات، خوشه‌بندی، کشف الگو و نظایر آن را شامل می‌شود. در واقع کارکرد اصلی این دسته از تحلیل‌ها آن است که ویژگی‌های غیر ضروری را حذف کرده و با فشرده کردن و تقلیل داده‌ها، امکان تحلیل بهتر داده‌ها را فراهم نمایند. تحلیل‌های تصمیم‌گیری نیز به کشف و استخراج قوانین تصمیم‌کمک می‌کنند. هر یک از رشته‌های دانشگاهی بر روی جنبه‌هایی خاص از تئوری مجموعه‌های راف تمرکز داشته‌اند. به‌عنوان مثال رشته ریاضیات بر روی تئوری، مفاهیم و توسعه مجموعه‌های راف تمرکز داشته و طراحی الگوریتم‌های این مجموعه‌ها و موارد مرتبط به آن در رشته کامپیوتر مورد بررسی قرار می‌گیرد. سایر رشته‌ها نیز بر روی کاربردهای مجموعه‌های راف متمرکز گردیده‌اند. به‌بیان دیگر، هر چند اساس این تئوری علوم کامپیوتر و ریاضیات است، اما در سایر علوم نظیر پزشکی، علوم اجتماعی و علم اقتصاد به روش‌های مختلفی از آن استفاده شده است (کریمی و صادقی، ۱۳۹۴).

موفقیت تئوری مجموعه‌های راف و توسعه آن را می‌توان به دلیل قابلیت‌ها و مزایای این تئوری دانست که مهم‌ترین آن‌ها به شرح زیر است (دیمیتراس و همکاران، ۱۹۹۹؛ گریکو و همکاران، ۱۹۹۸):
این تئوری برخلاف احتمال در آمار یا درجه عضویت در تئوری مجموعه فازی صرفاً بر اساس داده‌های اصلی بوده و نیاز به هیچ‌گونه اطلاعات خارجی ندارد (کروزینسکا و همکاران، ۱۹۹۲؛ دابویس و پرید، ۱۹۹۲؛ شورون و

۱.....
۲.....
۳.....

گزریمالاباس، ۱۹۹۳). مدل مجموعه‌های راف، نه تنها ابزاری مناسب برای تحلیل مشخصه‌های کمی، بلکه مناسب مشخصه‌های کیفی نیز می‌باشد. این تئوری، حقایق مهم مستر در داده‌ها را کشف و به زبان طبیعی قوانین تصمیم بیان می‌نماید. مجموعه قوانین تصمیم استخراج شده توسط مدل مجموعه راف، شرح کلی از دانش موجود در جداول اطلاعات مالی ارائه نموده و داده‌های اضافی که حاوی اطلاعات خاصی نمی‌باشند را حذف می‌نماید. قوانین به دست آمده از مدل مجموعه‌های راف مبتنی بر واقعیات می‌باشد، زیرا هر قانون تصمیم توسط یک سری نمونه‌های واقعی پشتیبانی می‌شود. نتایج مدل مجموعه راف، به راحتی قابل فهم می‌باشند، در حالی که نتایج حاصل از سایر روش‌ها، نیاز به تفسیر پارامترهای فنی داشته که ممکن است استفاده کننده با آن آشنا نباشد.

تئوری مجموعه‌های راف یک ابزار قدرتمند ریاضی به منظور مدیریت ابهام و عدم اطمینان ذاتی موجود در فرآیند تصمیم‌گیری می‌باشد.

مفاهیم تئوری مجموعه‌های راف بر این فرض متکی است که هر عضو مجموعه مورد بحث با برخی اطلاعات مرتبط است. به عنوان مثال، اگر اعضا، دارایی‌های موجود در یک بازار باشند، اطلاعات مربوط به دارایی‌ها شامل رفتار قیمت و ویژگی‌های اقتصادی می‌باشد. اعضایی که با اطلاعات مشابه مشخص می‌شوند، از منظر اطلاعات مستر در آن‌ها، غیر قابل تمایز (شبهه) می‌باشد. رابطه عدم تمایز، از این نظر، مبنای ریاضی در تئوری مجموعه‌های راف شده است. مهم‌ترین مسائل را که می‌توان با استفاده از تئوری مجموعه راف حل نمود عبارت‌اند از: تفسیر مجموعه‌ای از اعضا از نظر مقادیر مشخصه‌ها، کنترل وابستگی بین ویژگی‌ها (کامل یا جزئی)، کاهش مشخصه‌ها، تجزیه و تحلیل اهمیت ویژگی‌ها و تولید قوانین تصمیم‌گیری (پاولاک ۱۹۹۷).

مفاهیم بنیادی مجموعه‌های راف (فضاهای تقریب و تقریب مجموعه‌ها)

تئوری مجموعه راف یک روش ریاضی جدید برای دانش ناقص یعنی برای ابهام یا دقت پایین معرفی نمود. در این روش، ابهام توسط محدوده مرزی یک مجموعه بیان می‌شود.

حال، این مسئله را به طور دقیق تر تشریح می‌کنیم. فرض کنیم . یک مجموعه متناهی از اعضا بوده و یک رابطه زوجی داده شده باشد. مجموعه‌های . و . به ترتیب مجموعه مرجع و رابطه عدم تمایز نامیده می‌شوند. رابطه عدم تمایز، فقدان دانش ما را در مورد عناصر . نمایش می‌دهد. برای سهولت، فرض می‌کنیم که . یک رابطه هم ارزی می‌باشد. زوج . . . یک فضای تقریب نامیده می‌شود که ، مجموعه مرجع و . یک رابطه هم ارزی در . می‌باشد. فرض کنیم . یک زیرمجموعه . باشد، یعنی . هدف ما، توصیف مجموعه . بر اساس . می‌باشد. به منظور انجام این کار، به نشان گذاری بیشتر و مفاهیم پایه‌ای تئوری مجموعه راف نیاز داریم که در ذیل ارائه می‌شود.

با استفاده از ...، کلاس هم‌ارز را نشان می‌دهیم که توسط عضو تعیین می‌شود. رابطه هم‌ارزی، به یک معنا، فقدان دانش ما در خصوص مجموعه را شرح می‌دهد. کلاس‌های هم‌ارزی رابطه که قطعات^۱ نامیده می‌شوند، بخش‌های مقدماتی دانش است که ما قادریم با توجه به درک کنیم. به‌طور کلی، تنها با استفاده از رابطه عدم تمایز قادر به مشاهده اعضای جداگانه نیستیم بلکه صرفاً قطعات در دسترس دانش توسط این رابطه تشریح می‌شوند.

- مجموعه کلیه اعضای که می‌تواند با اطمینان به‌عنوان اعضای نسبت به طبقه‌بندی شوند، تقریب پایین مجموعه نسبت به^۲ نامیده می‌شوند و با ... نشان داده می‌شود، یعنی:

$$C \dots \dots \dots$$

- مجموعه کلیه اعضای که تنها می‌توان به‌عنوان اعضای محتمل نسبت به طبقه‌بندی نمود، تقریب بالای^۳ مجموعه نسبت به نامیده می‌شود و با ... نشان داده می‌شود، یعنی:

$$(X) = \{x \in X : R(x) \cap X \neq \emptyset\}$$

- مجموعه کلیه اعضای که نمی‌توان به‌طور قطع نه به‌عنوان عضو و نه به‌عنوان عضو نسبت به طبقه‌بندی نمود، محدوده مرزی^۴ مجموعه نسبت به نامیده می‌شود و با ... نمایش داده می‌شود، یعنی:

پژوهشگاه علوم انسانی و مطالعات فرهنگی

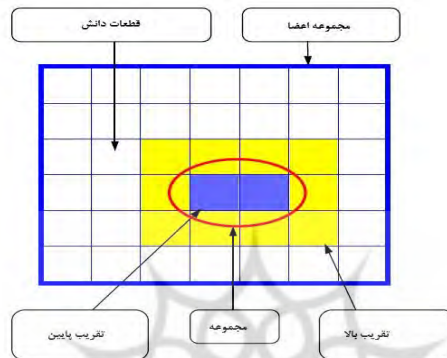
حال ما آماده‌ایم که علائم و تعاریف مجموعه‌های راف را فرمول‌بندی نماییم.

- یک مجموعه قطعی^۵ است اگر و تنها اگر محدوده مرزی تهی باشد.
- یک مجموعه نسبت به راف^۶ نامیده می‌شود اگر و تنها اگر محدوده مرزی تهی نباشد.

تعاریف مربوط به تقریب‌های مجموعه که در بالا ارائه شد را می‌توان بر اساس قطعات دانش به طریق زیر تعریف نمود. تقریب پایین یک مجموعه شامل اجتماع کل قطعاتی است که به‌طور کامل در مجموعه

1...
2...
3...
4...
5...
6...

قرار می‌گیرند؛ تقریب بالا، اجتماع کلیه قطعاتی است که تداخل غیر تهی^۱ با مجموعه دارند؛ محدوده مرزی یک مجموعه، تفاضل بین تقریب بالا و تقریب پایین مجموعه می‌باشد. شکل (۱) تصویر گرافیکی تقریب‌های تعریف شده در بالا را نشان می‌دهد.



شکل ۱. تصویر مفهومی تقریب‌های مجموعه

مقایسه تعاریف مجموعه‌های کلاسیک، فازی و راف جالب است. مجموعه کلاسیک یک مفهوم پایه‌ای و به صورت بصری و بدیهی تعریف می‌شود. مجموعه‌های فازی با استفاده از توابع عضویت فازی تعریف می‌شوند که دربرگیرنده ساختارها، اعداد و توابع ریاضی پیشرفته می‌باشند. توابع راف با استفاده از تقریب‌ها تعریف می‌شوند؛ بنابراین، این تعریف هم نیازمند مفاهیم پیشرفته ریاضی می‌باشد.

پیشینه تجربی پژوهش

کاربرد مدل مجموعه‌های راف در پیش‌بینی اقتصادی و مالی

کاربردهای مدل مجموعه‌های راف در پیش‌بینی اقتصادی و مالی را می‌توان به سه حوزه اصلی تقسیم نمود: پیش‌بینی شکست کسب‌وکار، بازاریابی پایگاه داده و سرمایه‌گذاری مالی. مراجع مربوطه به همراه مدل‌های مجموعه‌های راف مورد استفاده در جدول ۱ داده شده‌اند. کاربردهای شرح داده شده در اینجا نشان‌دهنده تنوع مسائلی است که با استفاده از مدل‌های مجموعه‌های راف قابل حل می‌باشند. در ادامه، جزئیات کاربرد مدل مجموعه راف در زمینه سرمایه‌گذاری مالی معرفی می‌شود:

1.....

جدول ۱. زمینه‌های اصلی کاربرد و مدل مجموعه‌های راف متناظر

مدل‌های مجموعه‌های راف	پیش‌بینی شکست کسب‌وکار	بازاریابی پایگاه داده‌ها	سرمایه‌گذاری مالی
.....			بازان و همکاران (۱۹۹۴) بالتزرسن (۱۹۹۶)
.....		پل (۱۹۹۸)	
.....	زیادو و میلز (۱۹۹۳)	میلز (۱۹۹۳) مروزک و اسکایک (۱۹۹۸)	زیارکو و همکاران (۱۹۹۳) گلان (۱۹۹۸) گلان و ادواردز (۱۹۹۳) روگریو (۱۹۹۴ ...) اسکالکو (۱۹۹۶) لین و ترمبا (۲۰۰۰)
.....		ایبن و همکاران (۱۹۹۸) کوالزیک و اسلیسر (۱۹۹۷) کوالزیک و پیاستا (۱۹۹۸) کوالزیک (۱۹۹۸)	
.....		پونل (۱۹۹۸) پونل و پیاستا (۱۹۹۸)	
.....	گریگو و همکاران (۱۹۹۸)		
.....	اسلوینسکی و زوپوندس (۱۹۹۴ و ۱۹۹۵) دیمیتراس و همکاران (۱۹۹۹) اسلوینسکی و همکاران (۱۹۹۷ و ۱۹۹۹)		ساسماگا و همکاران (۱۹۹۷)

سرمایه‌گذاری مالی

بسیاری از روش‌های تحلیل مالی از تکنیک‌های مدل‌سازی پیش‌بینی نظیر رگرسیون آماری و شبکه عصبی به‌منظور ایجاد و بهینه‌سازی پرتفوی‌ها و همچنین ایجاد سیستم‌های معاملاتی استفاده می‌کنند. ایجاد سیستم‌های معاملاتی با استفاده از مدل مجموعه‌های راف توسط پژوهشگران زیادی مطالعه شده است. زیارکو و همکاران (۱۹۹۳)، گلان و ادواردز (۱۹۹۳) و گلان (۱۹۹۵) به‌منظور کشف قوانین معاملاتی قوی، مدل مجموعه‌های راف را به کار گرفتند که الگوهای شدیداً تکراری در داده‌های استخراج‌شده از پایگاه داده تاریخی بورس اوراق بهادار تورنتو را منعکس می‌نمود. آن‌ها با استفاده از داده‌های اقتصادی و سهام سال ۱۹۸۵، قوانین معاملاتی سهام پنج شرکت بزرگ. . . را استخراج نمودند. این پنج شرکت عبارت‌اند از بانک مونترال، شرکت کانادایی بل، شرکت نفت امپریال، لابلز و شرکت ارتباطات نورثرن. این قوانین، رفتار سهام این شرکت‌ها و حساسیت آن‌ها به بازار و شاخص‌های اقتصادی را تشریح می‌نمود، اگرچه همه قوانین کشف توسط مدل مجموعه‌های راف حائز کیفیت بالا نبودند. قوانین عمومی (با پارامتر رافنس پایین^۱) که با استفاده از مدل مجموعه‌های راف استخراج شد همگی قوانین و روابط شناخته‌شده در صنعت سرمایه‌گذاری می‌باشند؛ درحالی‌که قوانین دقیق (با پارامتر رافنس بالا و نزدیک به ۱) سخت‌تری با آن‌ها (قوانین و روابط شناخته‌شده) دارد. کار آن‌ها یک متدولوژی بود که مدل مجموعه‌های راف را به‌عنوان یک گزینه کشف دانش در داده‌های بازار سهام تثبیت نمود. بازان و همکاران (۱۹۹۴) سیستم معاملاتی مشابهی را با استفاده از مدل مجموعه‌های راف بررسی نمودند. در این مطالعه، ۱۵ شاخص بازار جمع‌آوری شدند و نکته اصلی، تمرکز بر این موضوع بود که چگونه تعداد قوانینی که شاخص‌های مالی در انتهای یک ماه را در تغییرات قیمت سهم ماه بعد منعکس می‌نماید، کاهش می‌دهند. باهدف کاهش زمان محاسبه مجموعه بی‌زائده^۲، روش‌های حل جدید - خوشه‌بندی مفهومی، استخراج مشخصه‌های جدید از جدول تصمیم و ترکیب گروهی از مشخصه‌ها - موردبحث قرار گرفتند. نتایج اولیه استفاده از مدل مجموعه‌های راف تنها با دقت طبقه‌بندی ۴۴ درصد رضایت‌بخش بودند. بعلاوه، همچنان مشکلاتی نظیر فیلتر نمودن داده‌ها، داده‌های ناقص و مسئله یادگیری تکاملی وجود داشتند که باید مطالعه می‌شدند. بالتزرسن (۱۹۹۶) پژوهش مشابهی را با

۱. توضیح: قوانینی که توسط تعداد زیادی مثال پشتیبانی نمی‌شوند که ممکن است با تجربیات افراد خبره در تناقض باشند

(.)

استفاده از مدل مجموعه‌های راف بر روی شاخص کل بازار سهام اسلو انجام داد. مطالعات ایشان شامل جمع‌آوری داده‌ها، انتخاب و تغییر سری‌های زمانی به اعضای مجموعه‌های راف، آنالیز بی‌زائده و ایجاد قوانین با استفاده از مدل مجموعه‌های راف به‌منظور پیش‌بینی رشد شاخص بود. اگرچه دقت طبقه‌بندی رضایت‌بخش نبود (از ۲۵٪ تا ۴۵٪ متناسب با روش‌های مختلف گسسته‌سازی داده‌ها)، ولی وی چندین شاخص واضح را برای عواملی استخراج نمود که تغییرات آن‌ها تأثیر بیشتری نسبت به مقدار (سطح) آن‌ها دارد، نظیر شیب نرخ بهره، نرخ ارز که از خود این متغیرها مهم‌تر بودند. ایجاد سیستم معاملاتی بر روی شاخص یک کاربرد عمده دیگر در سرمایه‌گذاری مالی بود. اسکالکو (۱۹۹۶) با پیروی از روش‌های داده‌کاوی و کشف دانش، یک سری قوانین را از شاخص داده‌های آمار معاملات بورس سهام نیویورک. و داده‌های بازده اوراق قرضه خزانه آمریکا از اکتبر ۱۹۸۷ تا دسامبر ۱۹۹۴ استخراج نمود. این قوانین استخراج‌شده برای معاملات با استفاده از داده‌های ژانویه ۱۹۹۵ تا دسامبر ۱۹۹۵ به کار رفتند. با استفاده از این قوانین معاملاتی، ۹ معامله انجام شد از میان آن‌ها ۷ مورد سودمند بودند. مطالعه آن‌ها به دلیل استفاده از تحلیل تکنیکی و همچنین داده‌های مالی رفتاری، مورد توجه می‌باشد. روگریو (۱۹۹۴). پژوهش‌های زیادی به‌منظور ایجاد سیستم‌های معاملاتی بر روی شاخص. انجام داد. وی مجموعه‌ای از قوانین را برای پیش‌بینی موقعیت‌های خرید و فروش در. ایجاد نمود که چرخه‌های قیمتی مختلف بازار را تشخیص می‌داد. او ادعا نمود که عملکرد عالی با استفاده از قوانین قوی و کنار گذاشتن قوانین ضعیف به دست آمد. در طول کل دوره معامله، این سیستم در بیش از ۷۰٪ موارد در پیش‌بینی‌های پنج‌هفته آینده موفق بود و به‌طور میانگین بیش از ۲۵۰۰۰ دلار سود بر هر قرارداد. ارائه نمود. در پژوهش‌های ثانوی روگریو (۱۹۹۴)، قوانین استخراج‌شده، توان پیش‌بینی سودهای قوی در شاخص. با ۲٪ بازده بیشتر ارائه نمودند. این سیستم معاملاتی، قوانین عمومی قوی برای پیش‌بینی سیگنال‌های "خرید" و "فروش. استخراج نمود. روگریو (۱۹۹۴) همچنین یک سیستم معاملاتی ترکیبی ارائه نمود که شبکه‌های عصبی، مجموعه‌های راف و صفحات گسترده را ترکیب نمود. مدل‌های شبکه عصبی به‌منظور تصحیح خطاهای پیش‌بینی قوانین ایجادشده توسط مجموعه‌های راف بازنگری شدند. سیستم به کاربر اجازه می‌دهد تا ناکارآمدی موجود در بازار را رصد و از آن بهره‌برداری نماید. این سیستم معاملاتی به کاهش افت سرمایه از ۲۵٪ تا ۵۰٪ و افزایش میانگین نسبت برنده به بازنده از ۵۰٪ تا ۱۰۰٪ دست‌یافت. همچنین میانگین زمان معامله ۵۰ تا ۸۰٪ کاهش یافت.

مدل مجموعه‌های راف در حوزه مدیریت پرتفلیو نیز به کاررفته است. گریکو و همکاران (۱۹۹۶) برخی پژوهش‌ها را بر روی مسئله انتخاب سهم بازار سهام ایتالیا بر اساس روش مجموعه‌های راف انجام دادند. سهام بیست و دو شرکت موفق ایتالیایی که برای یک دوره طولانی فهرست شده بودند، به منظور انتخاب یک پرتفلیو با وزن برابر انتخاب شدند. هفت فاکتور به عنوان نماد ریسک برای تشکیل جدول تصمیم انتخاب شدند. پرتفلیو با استفاده از روش مجموعه‌های راف و دو تا از روش‌های سنتی مدل‌های چندعاملی مطالعه شد. نتایج نشان داد که در حالت نمونه‌های کوچک و دوره‌های پرنوسان، مجموعه‌های راف نسبت به روش‌های چندعاملی عملکرد بهتری دارند و یک ابزار نویدبخش به منظور مدیریت ریسک پرتفلیو با تمرکز بر مالی رفتاری می‌باشد. یکی دیگر از کاربردهای تئوری مجموعه‌های راف در می‌باشد. مفهوم در مدیریت پرتفلیو، تعریف یک روش سیستماتیک برای تشکیل پرتفلیو می‌باشد که دارای ارزشی بالاتر یا پایین‌تر نسبت به یک پرتفلیو معیار از نظر یک ویژگی خاص باشد (نظیر سازوکار میانگین واریانس). به عنوان مثال، یک سرمایه‌گذار پرداخت سود پایین سهم را بر پرداخت سود بالای سهم ترجیح می‌دهد و سایر موارد از نظر ایشان در اولویت یکسان می‌باشند. این موضوع بدین معنی است که پرتفلیو به گونه تشکیل شود که یا شامل تعداد بیشتری از سهام با پرداخت سود پایین شود و یا شامل سرمایه‌گذاری بیشتر در سهام با پرداخت سود پایین شود. ساسماگا و همکاران (۱۹۹۷) مسئله را بر اساس برخی مشخصه‌های کاملاً مرتبط با قیمت سهام مطالعه نمودند. مجموعه داده‌های استفاده شده شامل همه شرکت‌هایی می‌شود که به طور عمومی دادوستد شده و در فهرست شده‌اند. با استفاده از آزمایش اعتبارسنجی متقابل چندلایه مشخص شد که برای سرمایه‌گذاران بالقوه، مشخصه‌های مرتبط با قیمت در طراحی استراتژی‌های سرمایه‌گذاری موفق، دارای اهمیت بیشتری می‌باشند. این موضوع دلالت بر این دارد که تحلیل مجموعه‌های راف در تعیین نقش مشخصه‌ها در تشخیص پارامترهای مهم سهام مفید بوده و همچنین امکان استخراج قوانین تصمیم را می‌دهد که می‌تواند برای ارزیابی سهام جدید قابل کاربرد باشد. می‌توان گفت که مدل مجموعه‌های راف یک جایگزین نویدبخش برای روش‌های سنتی می‌باشد.

مدل پیشنهادی

بر اساس مفاهیم تعریف شده بالا، در این مطالعه مدلی پیشنهاد شده که چارچوب کلی آن در شکل ۲ نشان داده شده و جزئیات آن در ادامه تشریح می‌شود.

پیش‌پرداخت

در این مرحله دو کار به شرح ذیل انجام می‌گیرد:

تبدیل داده‌ها و انتخاب شاخص‌های تکنیکال اساسی

در فرآیند تبدیل داده‌ها، داده‌های قیمتی یک دوره زمانی (در اینجا به دلیل حذف اثرات ناشی از تقسیم سود، تاریخ بین دو مجمع متوالی) به‌عنوان مجموعه داده‌های آزمایشی سهام انتخاب شده‌اند. این مجموعه داده شامل پنج مقدار بنیادی روزانه مربوط به سهام (حداکثر قیمت، حداقل قیمت، قیمت بازگشایی، قیمت پایانی و حجم معامله سهام) و همین‌طور نوسان قیمت روزانه می‌باشد که بیان‌کننده تغییر قیمت سهام در هرروز مشخص نسبت به‌روز ماقبل می‌باشد. با استفاده از داده‌های مربوط به پنج مقدار بنیادی سهام، چندین شاخص تحلیل تکنیکی محاسبه می‌شوند که عبارت‌اند از میانگین متحرک (، . .)، مومنتم (، . .)، تصادفی (، . .)، تصادفی، شاخص قدرت نسبی (، . .) و . . ویلیامز.

معرفی شاخص‌های تکنیکال استفاده‌شده در مقاله

میانگین متحرک

در آمار، میانگین متحرک شاخصی است که به‌منظور تحلیل مجموعه‌ای داده‌ها با ایجاد یک سری از میانگین‌ها از زیرمجموعه‌های مجموعه اصلی استفاده می‌شود. در یک مجموعه از اعداد، برای یک زیرمجموعه با اندازه مشخص، با میانگین‌گیری از اولین زیرمجموعه، میانگین متحرک به دست می‌آید. سپس زیرمجموعه با اندازه مشخص را به جلو حرکت داده زیرمجموعه‌های جدیدی از اعداد را به دست می‌آوریم که از آن‌ها میانگین گرفته می‌شود. این فرآیند بر روی کل مجموعه داده‌ها تکرار می‌شود. میانگین متحرک معمولاً برای داده‌های سری زمانی استفاده می‌شود تا نوسانات کوتاه‌مدت را یکنواخت نموده و روندها و چرخه‌های طولانی‌مدت را برجسته نماید. آستانه بین کوتاه‌مدت و طولانی‌مدت بستگی به کاربردهای آن دارد و پارامترهای میانگین متحرک متناسب با آن تنظیم می‌شود. به‌عنوان مثال، میانگین متحرک اغلب در تحلیل تکنیکال داده‌های مالی نظیر قیمت‌های سهام و بازده و حجم معاملات استفاده می‌شود. میانگین متحرک دارای انواع مختلفی است که مهم‌ترین آن‌ها عبارت‌اند از میانگین متحرک ساده، میانگین متحرک تجمعی، میانگین متحرک وزنی، میانگین متحرک نمایی و ... است.

شاخص نرخ تغییر^۱

شاخص نرخ تغییر که از آن به عنوان مومنتم ساده نیز اشاره می شود، یک نوسانگر می باشد که درصد تغییر قیمت از یک دوره به دوره بعدی را اندازه گیری می کند. محاسبه ...، قیمت فعلی را باقیمت در . دوره قبل مقایسه می کند.

. قیمت پایانی . دوره قبل قیمت پایانی . دوره قبل - قیمت پایانی

شاخص قدرت نسبی^۲

این شاخص که توسط جی ولز و ایلدر^۳ تعریف شده، یک نوسان سنج مومنتم^۴ می باشد که سرعت و تغییر حرکات قیمتی را اندازه گیری می کند. . . . بین صفر و ۱۰۰ می تواند نوسان داشته باشد. به طور سنتی و بر اساس تعریف و ایلدر، زمانی که . . بالای ۷۰ باشد و زمانی که به پایین تر از ۳۰ برسد می باشد.

$$RSI = \dots \frac{\dots}{RS}$$

که

.. میانگین زیان میانگین سود

شاخص Williams%R

این شاخص توسط لری ویلیامز^۵ توسط فرمول زیر معرفی شد:

. حد اقل قیمت پایین - حداکثر قیمت بالا قیمت پایانی - حداکثر قیمت بالا

حد اقل قیمت پایین = حد اقل قیمت پایین برای دوره گذشته مورد نظر

حداکثر قیمت بالا = حداکثر قیمت بالا برای دوره گذشته مورد نظر

1.
2.
3.
4.
5.

دوره پیش‌فرض برای ۱۴ دوره‌ای می‌باشد که می‌تواند روز، هفته، ماه یا یک چارچوب زمانی روزانه باشد. این شاخص از صفر تا ۱۰۰- در نوسان است. مقادیر صفر تا ۲۰- به صورت و از ۸۰- تا ۱۰۰- در نظر گرفته می‌شود.

شاخص نوسانگر تصادفی^۱

این شاخص در حقیقت یک نوع مومنتم بوده که موقعیت قیمت پایانی را نسبت به دامنه حداکثر- حداقل در طول چند دوره مشخص می‌نماید. نوسانگر تصادفی از قیمت، حجم و موارد مشابه پیروی نمی‌نماید؛ بلکه از سرعت یا مومنتم قیمت پیروی می‌نماید. به‌عنوان یک قاعده، مومنتم قبل از قیمت تغییر جهت می‌دهد؛ بنابراین انحرافات صعودی یا نزولی در نوسانگر تصادفی می‌تواند به‌عنوان اخطار حرکت معکوس قیمت مورداستفاده قرار گیرد. نحوه محاسبه این شاخص بر اساس فرمول ذیل می‌باشد:

$$\dots \dots \dots \text{حداقل قیمت پایین} - \text{حداکثر قیمت بالا} \dots \dots \text{حداقل قیمت پایین} - \text{قیمت پایانی جاری} \dots \dots \dots$$
$$\dots = \text{میانگین متحرک ساده سه‌روزه} \dots \dots$$

در فرآیند انتخاب شاخص‌های تکنیکال اساسی، به‌منظور تعیین درجه ارتباط شاخص‌های تعریف‌شده با نوسانات روزانه قیمت، داده‌های مربوطه در قالب یک ماتریس همبستگی تحلیل می‌شوند و از میان همه شاخص‌ها، آن شاخص‌های اساسی که دارای درجه بالایی از همبستگی با نوسانات قیمت روزانه هستند انتخاب می‌شوند.

گسسته سازی داده‌ها

زمانی که هدف از تحلیل داده‌ها، کلاسه‌بندی و پیش‌بینی است، بدیهی است که درجه اهمیت هر یک از مشخصه‌ها باید مورد توجه قرار گیرد. در چنین حالت‌هایی معمولاً از اطلاعات کمکی در خصوص وزن هر یک از شاخص‌ها استفاده می‌شود؛ اما در رویکرد مجموعه‌های راف هیچ‌گونه اطلاعات اضافی تر از آنچه در جدول داده‌ها موجود است موردنیاز نیست. اصولاً تئوری راف سعی دارد تا از داده‌های موجود در سیستم اطلاعاتی، وزن و قدرت هر یک از مشخصه‌ها را در کلاسه‌بندی مشخص نماید؛ بنابراین از

1.....

برخی استراتژی‌ها برای طبقه‌بندی ارزش‌های واقعی استفاده می‌کند که کیفیت الگوریتم‌های یادگیری به این استراتژی‌ها وابسته است. این الگوریتم‌ها به دنبال یافتن برش‌هایی برای تعیین بازه‌ها هستند. تمامی مقادیر موجود در هر بازه به یک مقدار مشخص تبدیل می‌شوند و در نتیجه مشخصه‌های عددی به مشخصه‌هایی که می‌توانند دسته‌بندی شوند تبدیل می‌گردند. در این استراتژی‌ها از فرآیندهای تبدیل داده‌ها و یافتن برش‌های مناسب بر روی داده‌ها جهت تعیین بازه‌ها استفاده شده و به تمامی مقادیر درون یک بازه، ارزش برابر داده می‌شود. این کار باعث کاهش تعداد مقادیر یا ارزش‌های ممکن مشخصه‌ها شده و منجر به تولید قوانین مشخص‌تر و دقیق‌تر می‌شود.



شکل ۲. چارچوب مدل پیشنهادی

استخراج قوانین

استخراج قوانین با استفاده از تئوری مجموعه راف

قبل از تولید قوانین لازم است که نسبت به تولید بی زائده‌ها اقدام شود. یکی از مهم‌ترین جنبه‌ها در تحلیل جدول اطلاعاتی، کشف و حذف مشخصه‌های تکراری و نیز شناسایی مهم‌ترین مشخصه از درون داده‌ها است. مشخصه‌های تکراری به مشخصه‌هایی گفته می‌شود که حذف آن‌ها هیچ

تأثیری بر روی شدت وابستگی بین مشخصه‌های موجود و مشخصه‌های تصمیم ندارد. شدت وابستگی اندازه‌ای است که قدرت تمیز بین مشاهدات را نشان می‌دهد. حداقل زیرمجموعه مشخصه‌ها که شدت وابستگی آن حداکثر است، بی‌زائده نامیده می‌شود. محاسبه هسته و بی‌زائده یک جدول اطلاعاتی در واقع انتخاب مشخصه‌های مهم و مرتبط آن سیستم اطلاعاتی است. در تئوری راف، به حداقل مشخصه‌ها که قابلیت تمیز یک مشاهده از دیگر مشاهدات را در کلاس‌های تصمیم مختلف داشته باشد "بی‌زائده" گفته می‌شود. بی‌زائده ممکن است به صورت تقریبی بیان شود که قادر به تمیز درصد از مشاهدات بوده و به آن ... می‌گویند. محاسبه تمامی بی‌زائده‌ها یک مسئله است که روش‌های هیوریستیکی برای آن پیشنهاد شده است. برای به دست آوردن بی‌زائده می‌توان از تابع تمایز استفاده نمود که این تابع از ماتریس تمایز حاصل شده و روش به دست آوردن در قسمت‌های قبل شرح داده شد. لازم به ذکر است که می‌توان یک تابع تمایز را از همه ورودی‌های موجود در ماتریس تمایز ساخت که در این صورت بی‌زائده‌ی حاصل شده قادر به تمیز همه مشاهدات از همدیگر و با گروه‌های تصمیم مختلف است. همچنین می‌توان بی‌زائده‌هایی را به دست آورد که بر اساس تابع تمایز هر مشاهده ایجاد شده‌اند. قوانین حاصل از این بی‌زائده‌ها قدرت و دقت پیش‌بینی بالاتری دارند. هر بی‌زائده در مجموعه بی‌زائده‌های تولید شده دارای یک شاخص پشتیبان است. این شاخص نشان‌دهنده قدرت آن بی‌زائده بوده و برابر نسبت درست‌نمایی بی‌زائده ضربدر ۱۰۰ است. در این پژوهش، از میان روش‌های متعدد محاسبه بی‌زائده، از دو روش الگوریتم ژنتیک و الگوریتم جانشون استفاده شده است. تولید بی‌زائده‌ها با هدف استخراج قوانین تصمیم ایجاد می‌شود. برای کسب اطلاعات بیشتر در مورد این روش‌ها مراجعه شود به (لیانگ، ۲۰۰۹) و (استارزیک و همکاران، ۲۰۰۰)^۱. در سمت چپ قوانین تصمیم، ترکیبی از ارزش‌های مشخصه‌ها است که معمولاً مشخصه تصمیم تمامی مشاهداتی که با این ترکیب همخوانی دارند، برابر سمت راست قانون است. قوانینی که از بی‌زائده‌ها استخراج می‌شوند، می‌توانند برای کلاسه‌بندی داده‌ها به کار روند. مجموعه قوانین به دست آمده به عنوان یک ابزار کلاسه‌بندی به کاررفته و می‌توان از آن برای کلاسه‌بندی داده‌های جدید استفاده کرد. کیفیت قوانین به بی‌زائده‌های به دست آمده بستگی دارد. سعی بر آن است تا قوانینی که بخش بیشتری از مجموعه مشاهدات را پوشش می‌دهند تولید شوند. پوشش مجموعه مرجع با قوانین دارای عمومیت بیشتر، منجر به کوچک شدن مجموعه قوانین می‌شود.

بررسی قوانین استخراج شده

استفاده از برخی معیارهای کیفی در خصوص قوانین می‌تواند منجر به حذف برخی قوانین دارای کیفیت پایین گردد. قوانین با خواندن ارزش هر یک از مشخصه‌های موجود در بی زائده و ارتباط آن به یک یا چند گروه تصمیم ساخته می‌شوند. قسمت دوم یک قانون یا "بخش آنگاه" فقط شامل یک گروه تصمیم می‌شود مگر اینکه گروه تصمیم با توجه به مشخصه‌های موجود در بی زائده، راف باشد. برای ارزیابی قوانین از شاخص عمومیت^۱ قانون استفاده می‌شود که برابر درصد مشاهداتی است که ضمن تعلق به گروه تصمیم عنوان شده در "بخش آنگاه" شرایط "بخش اگر" را نیز دارا هستند. همچنین استفاده از شاخص دقت قانون که برابر درصد مشاهداتی است که شرایط "بخش اگر" را دارند و به گروه تصمیم "بخش آنگاه" تعلق دارند نیز به ارزیابی قوانین کمک می‌کند.

ارزیابی و مقایسه

ارزیابی و مقایسه (اعتبارسنجی) قوانین حاصل از اعمال روش‌های مختلف

با انجام مراحل قبل، به‌طور خلاصه فرایند شناسایی داده‌ها با تشخیص صحیح مسئله و سازمان‌دهی مشاهدات و مشخصه‌های مرتبط تکمیل می‌شود تا از این طریق خواسته‌ها مشخص گردد. پس از جمع‌آوری داده‌های موردنیاز برای هر یک از مشاهدات و مشخصه‌ها، لازم است تا بر روی داده‌ها پیش‌پردازش‌هایی نظیر کنترل سازگاری و یافتن خطاها، حذف نویزها و تکمیل داده‌ها صورت پذیرد تا اطمینان حاصل شود که مجموعه داده‌ها از حداکثر دقت برخوردارند و در نهایت نسبت به پردازش داده‌ها و استخراج قوانین اقدام نمود. در این مرحله داده‌ها به دودسته یادگیری و کنترل تقسیم می‌شوند. داده‌های گروه یادگیری جهت ساخت مدل از طریق ایجاد بی زائده و استخراج قوانین مورد استفاده قرار گرفته و داده‌های گروه کنترل برای آزمون اعتبار مدل به کار می‌روند. در این پایان‌نامه داده‌های اولیه با ۲۴ حالت شامل ترکیبات مختلف روش‌های گسسته سازی داده‌ها و روش‌های تولید بی زائده مورد تجزیه و تحلیل قرار گرفته و نتایج ارزیابی و باهم مقایسه شده‌اند.

استفاده از داده‌های سال‌های مختلف و مقایسه نتایج

در پایان، کلیه تحلیل‌های قسمت‌های قبل با استفاده از داده‌های سال‌های مختلف انجام و با توجه به میزان نوسان داده‌ها در هر سال کارایی مدل بررسی و نتایج باهم مقایسه شدند.

فرضیه پژوهش

مدل پیشنهادی نسبت به استراتژی خرید و نگهداری عملکرد بهتری دارد. نحوه آزمون فرض بدین ترتیب است که با اعمال قوانین استخراج شده بر روی داده‌های کنترل، نسبت به خرید، فروش یا نگهداری سهم در یک دوره زمانی اقدام می‌کنیم و بازده کل را با بازده حاصل از استراتژی خرید و نگهداری سهم در همان دوره مقایسه می‌کنیم.

روش شناسی پژوهش

روش پژوهش در این پژوهش کمی، توصیفی - کاربردی می‌باشد. داده‌های مورد بررسی شامل داده‌های سری زمانی مربوط به قیمت و حجم معاملات یک سهم خاص در بورس اوراق بهادار تهران می‌باشد. در این مطالعه داده‌های شش سال متوالی (۱۳۸۸/۰۵/۰۵ لغایت ۱۳۹۴/۰۴/۲۴) بانک صادرات مورداستفاده قرار گرفته است. حجم نمونه آماری به تعداد روزهای کاری بین تاریخ دو مجمع سالانه متوالی می‌باشد که برای سال‌های مختلف می‌تواند متفاوت باشد. برای سال مبنا که تحلیل‌های اولیه بر اساس داده‌های مربوط به آن انجام این بازه زمانی ۱۳۹۰/۰۵/۰۵ تا ۱۳۹۱/۰۴/۲۵ می‌باشد. داده‌های مورد نیاز با استفاده از سایت اینترنتی شرکت مدیریت فناوری بورس تهران استخراج شد و در این زمینه داده‌ها در دسترس می‌باشند.

تجزیه و تحلیل داده‌ها و آزمون فرضیه‌ها (مطالعه موردی: بانک صادرات)

به‌منظور تشریح دقیق‌تر مدل، در این بخش، نحوه استفاده از این مدل در مورد داده‌های بانک صادرات به‌عنوان مطالعه موردی، به‌صورت گام‌به‌گام بررسی می‌گردد. به دلیل تعدیل نوسان ناشی از تقسیم سود، داده‌های بازه زمانی بین دو مجمع عادی سالیانه که تقسیم سود را مصوب می‌کند مورداستفاده قرار می‌گیرند. با استفاده از داده‌های موردنظر، هر گام مطابق با سه مرحله معرفی شده در مدل پیشنهادی ((الف) پیش‌پردازش؛ (ب) استخراج قوانین؛ و (ج) ارزیابی و مقایسه) به شرح زیر انجام می‌گیرد:

گام اول: تبدیل داده‌ها و انتخاب شاخص‌های تکنیکال اساسی

در این گام، پنج مقدار بنیادی مربوط به داده‌های سهام بانک صادرات (قیمت بازگشایی، بالاترین قیمت، کمترین قیمت، قیمت پایانی و حجم معامله) از سایت اینترنتی شرکت مدیریت فناوری بورس تهران استخراج شده و در ادامه شاخص‌های تکنیکال معروف با استفاده از معادلات مربوطه محاسبه می‌شوند. به این ترتیب هفت شاخص تکنیکال محاسبه شدند و سپس شاخص‌هایی با

بالاترین درجه همبستگی با مشخصه تصمیم شامل میانگین متحرک، شاخص قدرت نسبی، . . . تصادفی، مومنتم و . . . ویلیامز به عنوان مشخصه های شرطی انتخاب شدند. از آنجا که هدف اصلی در این پژوهش، پیش بینی جهت حرکت قیمت سهم (افزایش، کاهش یا ثبات) در روز بعد به منظور اخذ تصمیم مناسب (خرید، فروش یا نگهداری سهم) می باشد، نوسانات قیمتی روزانه مربوط به روز بعد با استفاده از معادله زیر به عنوان شاخص تصمیم انتخاب شد.

(قیمت ابتدایی - قیمت انتهایی = نوسانات قیمتی روزانه روز بعد)

گام دوم: گسسته سازی داده ها

در این گام، گسسته سازی داده ها (تولید بازه های گسسته) با استفاده از تکنیک های مختلف معرفی شده در قسمت های قبلی انجام شده و سپس داده ها به دو گروه یادگیری و کنترل به منظور انجام مراحل بعدی تقسیم می شوند. بر اساس پیشنهاد کوزیاک^۱ مقدار ۰/۶۳۲ برای داده های گروه یادگیری و مقدار ۰/۳۶۸ برای داده های گروه کنترل مناسب است. داده های گروه یادگیری جهت ساخت مدل از طریق ایجاد بی زائده و استخراج قواعد مورد استفاده قرار می گیرند و داده های گروه کنترل برای آزمون اعتبار مدل به کار می روند.

گام سوم: استخراج قوانین با استفاده از تئوری مجموعه راف

بعد از مشخص نمودن جدول تصمیم که حاوی داده های سری زمانی مربوط به مؤلفه های شرطی و مشخصه تصمیم می باشد، جدول تصمیم به عنوان ورودی نرم افزار تعریف می شود. مراحل گسسته سازی، تولید بی زائده ها به روش های مختلف و استخراج قوانین توسط نرم افزار انجام می گیرد.

گام چهارم: بررسی قوانین استخراج شده

بررسی میزان دقت قوانین و تعداد داده هایی که هر کدام از قوانین را پشتیبانی می کنند در این مرحله انجام می گیرد.

گام پنجم: ارزیابی و مقایسه (اعتبارسنجی) قوانین حاصل از اعمال روش های مختلف

همان طور که قبلاً ذکر شد، با استفاده از ترکیب روش های مختلف گسسته سازی داده ها و تولید بی زائده ها، در مجموع، داده های گروه یادگیری به ۲۴ حالت مورد تجزیه و تحلیل قرار گرفتند و قوانین استخراج شده در هر حالت با استفاده از داده های گروه کنترل مورد آزمون قرار گرفته و شاخص دقت

روش‌های مختلف باهم مقایسه شدند. هرچه شاخص دقت بالاتر باشد، قدرت پیش‌بینی قوانین حاصل از پردازش داده‌های یادگیری بالاتر می‌باشد. به‌عنوان مثال در شکل ۳، از ۶۴ داده کنترل، ۱۹ مورد صحیح پیش‌بینی افزایش قیمت، ۷ مورد صحیح پیش‌بینی ثبات قیمت، ۳۱ مورد پیش‌بینی صحیح کاهش قیمت یعنی به ترتیب ۹۵٪، ۶۴٪ و ۱۰۰٪ پیش‌بینی صحیح وجود داشته که دقت کلی روش در مجموع ۸۹٪ می‌باشد. در همین جا می‌توانیم فرضیه مطرح‌شده برای سنجش میزان کارایی مدل را با استفاده از داده‌های کنترل بررسی نماییم. نحوه آزمون فرض بدین ترتیب است که با اعمال قوانین استخراج‌شده بر روی داده‌های کنترل، نسبت به خرید، فروش یا نگهداری سهم در یک دوره زمانی اقدام می‌کنیم و بازده کل را با بازده حاصل از استراتژی خرید و نگهداری سهم در همان دوره مقایسه می‌کنیم. با توجه به اینکه پیش‌بینی انجام‌شده توسط قوانین در ۸۹٪ موارد صحیح بود، بازده حاصل از اعمال مدل به ازای هر سهم ۱۸۲ ریال در مقابل بازده ۳۳ ریال حاصل از استفاده از استراتژی خرید و نگهداری می‌باشد.

		Predicted				
		U	N	L	Undefined	
Actual	U	19	2	0	2	0.826087
	N	1	7	0	0	0.875
	L	0	2	31	0	0.939394
	Undefined	0	0	0	0	Undefined
		0.95	0.636364	1.0	0.0	0.890625
ROC	Class	Undefined				
	Area	3.402820e+038				
	Std. error	3.402820e+038				
	Thr. (0, 1)	3.402820e+038				
	Thr. acc.	3.402820e+038				

شکل ۳. نتایج اعتبار سنجی مدل

گام ششم: استفاده از داده‌های مربوط به سال‌های مختلف در مدل و مقایسه نتایج کلیه اقدامات انجام‌شده در گام‌های قبل که بر روی داده‌های مربوط به سال ۱۳۹۱-۱۳۹۰ انجام شد که در مجموع ۲۴ روش تحلیل می‌باشد، بر مبنای داده‌های پنج سال دیگر نیز اجرا شد که نتایج کلی در شکل‌های ۴ و ۵ جمع‌بندی شده است.

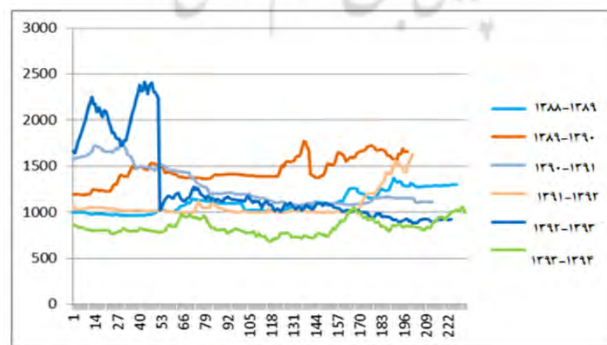
یکی از مزایایی که در قسمت‌های پیشین در مورد روش پیشنهادی ذکر شد، عدم نیاز به مفروضات خاص - برخلاف سایر روش‌ها - می‌باشد که در اینجا این موضوع نیز تأیید شد. همان‌طور که در متدولوژی ذکر شد، در بین شاخص‌های تکنیکال (متغیرهای شرطی)، شاخص‌هایی که بیشترین همبستگی را با متغیر تصمیم داشتند انتخاب‌شده و تحلیل‌ها بر این اساس انجام گرفت که نتایج قابل قبولی به دست آمد. در ادامه تحلیل‌های مربوط به

داده‌های سال‌های مختلف با میزان همبستگی‌های متفاوت بین مؤلفه‌های ذکر شده، بازهم نتایج قابل قبول و در برخی موارد نتایج بهتری حاصل شد که تا حد زیادی این ویژگی مهم تئوری مجموعه‌های راف یعنی عدم نیاز به مفروضات اولیه را تأیید می‌نماید. همان‌طور که در جدول ۲ مشخص است در مورد داده‌های سال ۱۳۹۱-۱۳۹۰ که هر پنج مشخصه شرطی با مشخصه تصمیم همبستگی دارند شاخص اعتبارسنجی مدل ۰/۸۹۰۶ است، در حالی که این شاخص برای سال ۱۳۹۲-۱۳۹۱ که مشخصه‌های فوق‌الذکر باهم همبستگی در سطح بالایی ندارند، ۰/۹۱۰۴ می‌باشد. به عبارت دیگر، در این مطالعه، همبستگی بین داده‌های مربوط به مشخصه‌های شرطی و مشخصه تصمیم تأثیر معنی‌داری در نتایج نهایی بروز نداده است.

جدول ۲. مقایسه معیار سنجش قدرت قوانین در سال‌های مختلف (یافته‌های پژوهش)

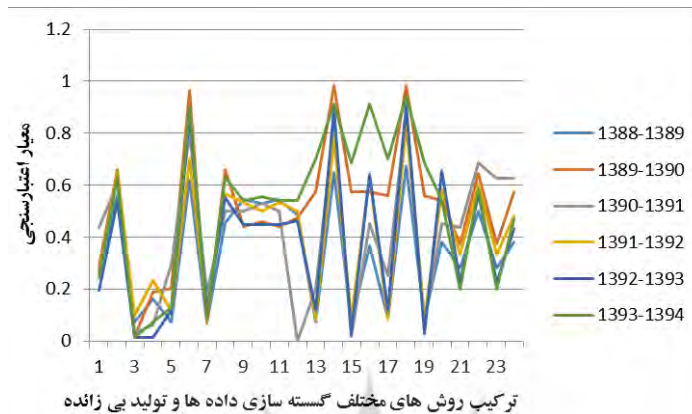
معیار سنجش قدرت قوانین	متغیرهای شرطی					سال
	RSI	R	MTM	MA	K	
۰.۶۷۶۴	-	*	-	-	*	۲۰۰۹-۲۰۱۰
۰.۹۸۳۰	-	*	*	*	*	۲۰۱۰-۲۰۱۱
۰.۸۹۰۶	*	*	*	*	*	۲۰۱۱-۲۰۱۲
۰.۸۳۳۲	*	-	*	-	-	۲۰۱۲-۲۰۱۳
۰.۹۱۰۴	-	-	-	-	-	۲۰۱۳-۲۰۱۴
۰.۹۴۲۸	-	*	-	*	*	۲۰۱۴-۲۰۱۵

توضیح: متغیرهایی که با علامت (*) مشخص شده‌اند، بر اساس داده‌های هر سال دارای همبستگی بالایی با متغیر تصمیم می‌باشند.



زمان (روزهای معاملات سهم در سال)

شکل ۴. نمودار مقایسه نوسان قیمت سهم در سال‌های مختلف



شکل ۵. نمودار مقایسه اعتبارسنجی روش‌های متفاوت در سال‌های مختلف (یافته‌های پژوهش)

نتیجه‌گیری و بحث

تئوری مجموعه‌های راف برخلاف احتمال در آمار یا درجه عضویت در تئوری مجموعه فازی صرفاً بر اساس داده‌های اصلی بوده و نیاز به هیچ‌گونه اطلاعات خارجی ندارد. در این مطالعه نیز صرفاً از داده‌های موجود بدون لحاظ مفروضات خاصی استفاده شده است. روش‌های متعددی جهت ترمیم داده‌ها در صورت نیاز وجود دارد که استفاده از این روش‌ها در سایر زمینه‌های کاربرد مجموعه‌های راف با داده‌های ناقص مفید می‌باشد. اثبات غیر نرمال بودن داده‌ها و همچنین عدم همبستگی برخی شاخص‌های تکنیکال در سطح اطمینان قابل قبول با مشخصه تصمیم (با استفاده از آزمون‌های آماری) منجر به تأثیر غیرقابل انتظار یا مخدوش شدن نتایج به دست آمده نگردید که می‌توان گفت به نوعی مهر تأیید بر ادعای فوق می‌باشد. مدل مجموعه‌های راف، نه تنها ابزاری مناسب برای تحلیل مشخصه‌های کمی، بلکه مناسب مشخصه‌های کیفی نیز می‌باشد. به عنوان مثال، در این مطالعه علاوه بر مشخصه‌های شرطی که به صورت کمی بودند، مشخصه تصمیم که در ابتدا به صورت کمی بود، به سه حالت کیفی یعنی افزایش قیمت در روز بعد (.)، کاهش قیمت در روز بعد (.) و عدم تغییر (.) تبدیل شد و نتایج مفیدی نیز به دست آمد. این تئوری، حقایق مهم مستتر در داده‌ها را کشف و به زبان طبیعی قوانین تصمیم بیان می‌نماید. خروجی نهایی تحلیل‌های انجام شده، یک سری قوانین ساده، قابل فهم و کاربردی است. مجموعه قوانین تصمیم استخراج شده توسط مدل مجموعه راف، شرح کلی از دانش موجود در جداول اطلاعات مالی ارائه نموده و داده‌های اضافی که حاوی اطلاعات خاصی نمی‌باشند را حذف می‌نماید. قوانین به دست آمده از مدل مجموعه‌های راف مبتنی بر واقعیات می‌باشد، زیرا هر قانون تصمیم توسط یک سری

نمونه‌های واقعی پشتیبانی می‌شود. نتایج مدل مجموعه راف، به‌راحتی قابل فهم می‌باشند، درحالی که نتایج حاصل از سایر روش‌ها، نیاز به تفسیر پارامترهای فنی داشته که ممکن است استفاده کننده با آن آشنا نباشد. تئوری مجموعه‌های راف یک ابزار قدرتمند ریاضی به‌منظور مدیریت ابهام و عدم اطمینان ذاتی موجود در فرآیند تصمیم‌گیری می‌باشد. مزیت اصلی مدل این است که در هر صورت، چه بازار صعودی باشد و چه نزولی، بازده مثبت دارد. در بازار صعودی، بازده هر دو روش یعنی مدل پیشنهادی و استراتژی خرید و نگهداری مثبت است؛ لکن، مدل پیشنهادی بازده بالاتری ارائه می‌دهد. در بازار نزولی که بازده حاصل از استراتژی خرید و نگهداری منفی است، بازده ناشی از به‌کارگیری مدل مثبت است که این شواهد نشان‌دهنده توانایی استثنایی مدل پیشنهادی در استخراج الگوهای قیمتی صحیح در بازار سهام می‌باشد. مقایسه نتایج حاصل از اعمال روش‌ها بر روی داده‌های مربوط به سال ۱۳۹۱-۱۳۹۰ نشان می‌دهد که بازده حاصل از استراتژی خرید و نگهداری ۳۳ ریال و بازده حاصل از مدل ۱۸۲ ریال به ازای هر سهم می‌باشد. در این مطالعه، با ترکیب روش‌های مختلف گسسته‌سازی داده‌ها و روش‌های تولید بی‌زائده، در مجموع داده‌های مربوط به هر سال به ۲۴ حالت مورد تجزیه و تحلیل قرار گرفتند که نتایج حاصل از روش الگوریتم گسسته‌سازی ساده و روش تولید بی‌زائده الگوریتم ژنتیک بهترین شاخص اعتبارسنجی را ارائه نمودند. نکته مهم و قابل توجه، همخوانی نتایج کاربرد روش‌های مختلف بر روی داده‌های سال‌های متفاوت (شش سال متوالی) می‌باشد؛ به‌طوری که نمودار نتایج حاصل از روش‌های مختلف در سال‌های مختلف همبستگی جالب توجهی دارد و حداکثر بازده در سال‌های مختلف ناشی از اعمال روش ذکر شده در بالا می‌باشد. شکل ۵ گویای این موضوع می‌باشد. فرآیند گسسته‌سازی داده‌ها یک روش پیش‌پردازش چندمنظوره است که تعداد مقادیر مجزای یک متغیر مشخص را با تقسیم محدوده آن به مجموعه‌ای محدود از بازه‌های تفکیک‌شده کاهش می‌دهد و سپس این بازه‌ها را به شکل معنی‌داری به هم مرتبط می‌نماید. در ادامه، داده‌ها، به‌جای مقادیر مجزای دقیق، آنالیز شده یا در سطح بالاتری از دانش نمایش داده می‌شوند و بنابراین منجر به نمایش ساده‌ای از داده‌ها در داده‌کاوی و اکتشاف داده‌ها می‌شود. همان‌طور که قبلاً نیز ذکر شد از امتیازات گسسته‌سازی داده می‌توان به نیاز به حافظه کمتر برای مشخصه‌های گسسته، نزدیک‌تر بودن به نمایش در سطح دانش، ساده‌سازی و کاهش داده‌ها و در نتیجه استفاده و توصیف آسان‌تر و یادگیری دقیق‌تر و وسیع‌تر با استفاده از مشخصه‌های گسسته اشاره نمود. البته آنچه در این مطالعه اثبات شد، کارایی متفاوت روش‌های مختلف تولید بازه‌های گسسته است. یکی از مهم‌ترین جنبه‌ها در تحلیل جدول اطلاعاتی، کشف و حذف مشخصه‌های تکراری و نیز شناسایی مهم‌ترین مشخصه‌ها از درون داده‌ها است. مشخصه‌های تکراری به مشخصه‌هایی گفته می‌شود که حذف آن‌ها هیچ تأثیری بر روی شدت وابستگی بین مشخصه‌های موجود و مشخصه‌های تصمیم‌ناردار. فرآیند حذف مشخصه‌های تکراری از طریق الگوریتم

ژنتیک و الگوریتم جانسون (هر کدام به دو حالت) انجام گرفت که در مجموع نتایج حاصل از الگوریتم ژنتیک دقیق‌تر بودند. از ترکیب روش‌های گوناگون گسسته سازی داده‌ها و تولید بی‌زائده‌ها، در مجموع ترکیب روش گسسته سازی ساده با روش تولید بی‌زائده الگوریتم ژنتیک منجر به نتایج بهتری شدند. در این مطالعه، ارتباط معنی‌داری بین میزان نوسانات یا روند کلی قیمت سهم در یک دوره خاص با میزان قدرت پیش‌بینی مدل مورد استفاده پیدا نشد. به عبارت دیگر، اگرچه در مجموع نتایج حاصل از تحلیل داده‌های سال‌های مختلف رضایت‌بخش بودند، لکن، به عنوان مثال دو سال متوالی ۱۳۸۸-۱۳۸۹ و ۱۳۹۰-۱۳۸۹ با روند و الگوی نوسانی تقریباً مشابه، به ترتیب حائز کمترین و بیشترین میزان شاخص اعتبارسنجی شدند.

به منظور انجام پژوهش‌های آتی، پیشنهاد می‌گردد از سایر روش‌های گسسته سازی و تولید بی‌زائده استفاده شود. همچنین پیشنهاد می‌شود از ترکیب روش مجموعه‌های راف با سایر روش‌ها نظیر تئوری فازی و الگوریتم ژنتیک استفاده شود. در این مطالعه، جدول تصمیم به گونه‌ای طراحی شده بود که پیش‌بینی‌ها در هر روز برای روز بعد امکان‌پذیر بود؛ بر همین اساس پیشنهاد می‌شود مشخصه‌های شرطی و مشخصه تصمیم به نحوی تعریف شوند که بتوان به صورت پیوسته در طول یک روز نیز بتوان از نوسان قیمت سهم استفاده کرده و سود به دست آورد. پیشنهادها دیگری که با محوریت متدولوژی این پایان‌نامه می‌توان ارائه داد، یکی استفاده از شاخص‌های تکنیکالی که در اینجا استفاده نشده‌اند و دیگری بررسی روش‌هایی برای حذف نویزهایی از قبیل تقسیم سود می‌باشد. در پایان پیشنهاد می‌شود یک برنامه کامپیوتری طراحی شود که با استفاده از قوانین استخراج شده و مقایسه آن‌ها با شرایط موجود، پیش‌بینی را انجام دهد. این نوع کد نویسی به‌طور خاص برای پیش‌بینی‌های کوتاه‌مدت و در حالتی که تعداد قوانین زیاد باشد، بسیار کاربردی است.

منابع

- فراهانی فرد. کامیار، قاسمیان لنگرودی. رضا، ۱۳۹۱، تحلیل تکنیکال در بازار سرمایه، آگاه، تهران، چ هشتم.
- کریمی. تورج، صادقی مقدم. محمدرضا، ۱۳۹۴، مجموعه‌های راف و مجموعه‌های خاکستری (مبانی، کاربرد، نرم‌افزار)، کتاب مهربان، تهران، چ دوم.
- Bll tzzree,, J. K. (1996). nnn attmmtt eets ppprccch” Dectrral disserttti,,, Kwwwlggge yyttmns Gr,,, Tee ..
- Bzz,,, J,, kkwwr,,, A,, & yykkk, .. (1994). MMrket ttt a nll yii: A rggge eet ppprcc...
- Cgggg, C. H,, C,,, T. ,, & ee i, .. Y. (2010). AAyyrr id mllll sss ed nn rogh eets toory ddd geetic ll grr itmns for ttokk rree forcsstigg..
- .. iiii nsss fiilrre rrdictiii iii gg rgggh eet” uuraaan Jrr nll of ..
- .. uuuttigg rgggh stts ddd fzyy eets tggteerII Itt ll liggth Dcciiioooooorto 203.....
- .. irrk,, .. (1995). AAmttlllll oyy fr ttokk mrrkt ll lyis utilizigg rgggh set trrr y.. In Cmmtt tionll Itt ll ligcce for ..
- Gll,,, R,, & wwvr,,, D. (1994). “mmrrr ll rulss iisvvrrii iii ggc/R+ with stkkk marktt tt ””. In ccceei ggs ff the Itt rraatinnll ..
- Grcc,, ,, Csci,, .. ,, & aa trzzz,, B. (1996). ooggheet ppr och to ttkkk eelcti:: An pppliaation to tee Itl inn mrrkt.. eeee ll... ..

- Grcc,, ,, aa trrzzz,, B,, & iiiii iii i, R. (1998). AA wwwrgggh stt rrrr occh to vvll ttt inn of kkkkrttt yy rikk. Orrr ttiaaal tools in tee
- łwwikkki, R. (1999). llllll igg missigg rrrlll mm” Itt rrrttt iaaal rrr kppppp pp Roggh ttt s, zzzyy ttt ,, Dtt a
- Kim, K. J,, & H,, I. (2000). eeeee tic algrr itmns aeeeahh to faature ixxxee eeeert eettmm ttt t t ll iaatinn,, 19(2), 125.... ..
- Kim, .. J,, ii ,, .. H,, & H,, I. (2006). mnn vvll utinnrry rrrr occh to the mmmi ttt inn of mll till e llssii firrs to prddict a stkkk rr ice ieee... Epp....
- rkk,, ,, wwwikkki, R,, & eefkkkkkki, kk(1992). DDccriminnt vrrsss rgggh eets rrrr ccch to vggee aata aaalyii”. Alll idd Sthhsstic
- aaagg, ,, (2009). “. rrrr att inn: The Appliaation ff tee Clssterigg.. ii fth Itt rraatinnll Joint
- ,, T,, & eem,, J. (2000). AAtribtt e Trsss fir mtt inns nn Nmnrriaal Dttaaees.. Kwwvlggge Diccvrry ddd
- Nikll lllll ll C,, & ll lrtt,, .. (1994). AA yyr id eeeert yyttmmfir ivvsstmttt vvvisiggee eeeert eettmm, 11(4), 245.... ..
- rrr k, C. H,, & Irwi,, .. H. (2007). WWtt oo we kwwwtttt teety ff tchii ell nll yii??. rrrr nll ff Eooomic vvvvy,, 21(4),
- wwvkk, Z. (1982). gggggh eet”. Itt rraatioggg Juurlll of Cmmtt er &
- wwvkk, Z. (2002). gggggh eets agg intlil lignnt ttt a aaalyii”. Iff rr mtt inn
- wwvkk, Z,, & Skww,, A. (2007). gggggh eets nnd Bgggan ressnigg..

- wkk, Z., & Skwr, A. (2007). mmmmmmmmm of rgggh eets.. Iffrr mtt inn
-
- R,, T. H. (2007). rrrr ccastigg the vll tt ility ff tttkkk rr ice in"""" . eeee rt
- Rgggirr, .. (1994). uuules rre mee to ee tree"" . AI in aaaacc,, 35... ..
- Rgggirr, .. (1994). wwwwo iii ld a yystmmfrmnr k. rrrrr ss, 23(12),
-
- kkw, A. (1993). llll aan raasll gg fir ccisinn rll ss geerati... In
- aarzyk, J. A., Nel,,, D. ,, & Sturtz, K. (2000). AA aa tmmtticllon in Iffrr mtt inn yyttmm""
- aaaaj, Z. (2004). nnn intrddttt inn to rgggh eet thrrr y add its alll iaati""""
-). Itttt ifyigg rggll rrities in stkkk rrr tfolio tiltigg,, Itt rrtt iaal Itt itute fir Applied
- Tyy, .. ,, & ,,, , (2002). EEmmmmmmmrammfiaaccill rrii tt inn uii gg rgggh eets mllll .. uurppaan rrrr ll ff Oeertt innal Rssaar,, 141(3... .. .
- Zirrk,, ,, Gol,, R,, & wwwr,, D. (1993). nnn pppliaatinn of tttkkk mrrktt ttt "" . In rreeddiggs ff AAAI wrksprr nn kwwlegge