پژوهش‌های نوین در تصمیم‌گیری
دوره ۷، شماره ۲، تابستان ۱۴۰۱، صص ۱۳۷-۱۶۷
نوع مقاله: پژوهشی

پیش‌بینی قیمت سهام با ارائه مدلی ترکیبی با استفاده از تحلیل مؤلفه‌های اصلی و تئوری مجموعه‌های راف

محمدرضا مهربان‌پور[۱]، عادل آذر[۲]*، مجید شهرامی بابکان[۳]

۱- استادیار، گروه حسابداری، دانشکده مدیریت و حسابداری، دانشکدگان فارابی، دانشگاه تهران، قم، ایران
۲- استاد، گروه مدیریت صنعتی، دانشکده مدیریت و حسابداری، دانشگاه تربیت مدرس، تهران، ایران
۳- دانشجوی دکتری مالی گرایش بانکداری، دانشکده مدیریت و حسابداری، دانشکدگان فارابی، دانشگاه تهران، قم، ایران

تاریخ دریافت: ۱۴۰۰/۰۸/۱۹ تاریخ پذیرش: ۱۴۰۱/۰۱/۱۹

چکیده

در این پژوهش، با ترکیب روش‌های تحلیل مؤلفه‌های اصلی و مجموعه‌های راف، مدلی به منظور پیش‌بینی قیمت سهام ارائه شده است. به این منظور، با استفاده از داده‌های قیمتی شرکت ایران خودرو، ابتدا تعدادی از شاخص‌های تکنیکال محاسبه شدند. به منظور کاهش بعد ماتریس تصمیم، به روش تحلیل مؤلفه‌های اصلی، متغیرهایی جدید به گونه‌ای انتخاب شدند که حداکثر ویژگی‌های داده‌های اولیه حفظ شود. از این متغیرها در ماتریس تصمیم، به عنوان مؤلفه‌های شرطی استفاده می‌شود و متغیر تصمیم، نوسان قیمت سهم در روز بعد می‌باشد. داده‌ها به روش‌های مختلف گسسته‌سازی و به دو دسته یادگیری و کنترل تقسیم شدند. سپس با استفاده از تئوری مجموعه‌های راف بر روی داده‌های یادگیری، قواعد تصمیم استخراج و اعتبار آنها بر روی داده‌های کنترل، ارزیابی شد. نتایج به دست آمده از مدل ترکیبی با نتایج حاصل از مدل مجموعه‌های راف مقایسه شد. با بررسی ضرایب متغیرهای اولیه در عامل‌های جدید این موضوع مهم قابل نتیجه‌گیری است که با حفظ بخش عمده خواص داده‌های اولیه، پنج متغیر اولیه قابل کاهش به دو عامل جدید بوده و قابلیت نام‌گذاری دارند که این موضوع نقش مهمی در کاهش تعداد قواعد تصمیم و ملموس بودن استفاده از آنها دارد. درصد پیش‌بینی‌های صحیح قواعد استخراج شده از مدل ترکیبی نسبت به مدل رقیب یعنی مدل مجموعه‌های راف، بیشتر و تعداد قواعد کمتر می‌باشد. به منظور بررسی استحکام مدل، داده‌های بازه سال‌های ۱۳۹۸-۱۳۸۳ شرکت ایران خودرو و همچنین داده‌های سالیانه بانک صادرات ایران مورد بررسی قرار گرفتند که نتایج با یافته‌های قبلی مطابقت داشتند.

کلیدواژه‌ها: پیش‌بینی قیمت سهام، تحلیل مؤلفه‌های اصلی، قواعد تصمیم، مجموعه‌های راف

---

* نویسنده مسئول مقاله E-mail: azara@modares.ac.ir

۱- مقدمه

یکی از مهم‌ترین جنبه‌های قدرت انسان در کنترل و استفاده مطلوب از پدیده‌ها، توان پیش‌بینی رفتار آنها می‌باشد. نقش پیش‌بینی در پیشرفت‌های خیره‌کننده انسانی در علوم فیزیکی آشکار است. پیش‌بینی حرکت اجرام آسمانی و تغییرات جوّی از جمله این مسائل می‌باشد. هنگامی که موضوع پیش‌بینی، پدیده‌هایی است که انسان در آن نقش دارد، سوگیری‌های رفتاری، اعتبار پیش‌بینی را با مخاطراتی روبرو می‌سازد. پیچیدگی موضوعات وابسته به رفتار انسان‌ها تا حدی است که دانشمندان علوم انسانی گفته‌اند: "خداوند مسائل آسان را در مقابل فیزیک‌دانان قرار داده است"[۱]. تصمیم‌گیری در دنیای پیچیده امروز، به چالشی برای مدیران و سازمان‌ها تبدیل شده است. تعدد شاخص‌های تصمیم‌گیری، تنوع معیارهای کمی و کیفی و لزوم در نظر گرفتن آنها، اهمیت آثار و پیامدهای تصمیم و عواملی نظیر آن بر پیچیدگی تصمیم‌ها می‌افزاید [۲]. در بحث سرمایه‌گذاری مالی، پیش‌بینی قیمت سهام مقوله‌ای بسیار مهم است. تلاش‌های بسیاری در این راستا، با به‌کارگیری تکنیک‌های عاریت گرفته از علوم مختلف نظیر آمار، ریاضیات و فیزیک انجام گرفته است.

از نظر پیش‌بینی پذیربودن یا نبودن بازار، محققان و دانشگاهیان به دو گروه تقسیم می‌شوند؛ گروهی به کارآیی اطلاعاتی بازار اعتقاد دارند و براین باور هستند که بازار سهام، از یک گام تصادفی پیروی می‌نماید و دیگر جایی برای پیش‌بینی باقی نمی‌ماند (برای مثال: [۳]؛[٤]). گروه دوم سازوکارهایی را برای پیش‌بینی ارائه می‌دهند (مانند: [٥] ؛ [٦]؛ [۷]).

در راستای نظر گروهی که بازار را قابل پیش‌بینی می‌دانند، روش‌های متعددی به‌منظور پیش‌بینی قیمت سهام استفاده شده‌اند که به‌طور کلی می‌توان این روش‌ها را به چهار دسته طبقه‌بندی نمود: ۱) روش‌های تحلیل تکنیکی ۲) روش‌های تحلیل بنیادی ۳) روش‌های پیش‌بینی سری زمانی سنتی و ۴) روش‌های یادگیری ماشینی. پیشرفت‌های اخیر در تجزیه‌وتحلیل سهام و پیش‌بینی شامل چهار دسته آماری، شناخت الگو، یادگیری ماشینی[۱] و تحلیل احساسات، عمدتاً در زیرمجموعه روش تحلیل تکنیکی قرار داده شده‌اند؛ با این حال، برخی از تکنیک‌های یادگیری ماشینی وجود دارد که روش‌های تجزیه‌وتحلیل تکنیکی را با رویکردهای تجزیه‌وتحلیل بنیادی ترکیب می‌نماید [۸].

هرکدام از روش‌ها، متناسب با پیش‌فرض‌ها و نیز قدرت تکنیک‌های مورد استفاده، دارای نقاط قوت و ضعف خاص خود می‌باشند. بنابراین بهتر است تحلیل‌گر به جای اصرار بر انتخاب یک روش به‌عنوان روش برتر، با همه آنها آشنایی داشته باشد و حتی‌الامکان از مزایای

روش‌های مختلف و ترکیبی بهره‌برداری نماید. در سال‌های اخیر، با توجه به افزایش قابل توجه حجم داده‌ها، کاربرد روش‌های مبتنی بر داده‌کاوی نظیر شبکه‌های عصبی مصنوعی، منطق فازی، الگوریتم ژنتیک و ... به منظورهای مختلف نظیر طبقه‌بندی و پیش‌بینی افزایش چشم‌گیری یافته است.

امروزه بازار سهام، برخلاف سیستم‌های سنتی، در برگیرنده ترکیبی از فناوری‌های مختلف مانند یادگیری ماشینی، سیستم‌های خبره و داده‌های بزرگ بوده که با هدف تسهیل تصمیمات آگاهانه‌تر با یکدیگر در ارتباط می‌باشند. علاوه بر این، ارتباط جهانی کاربران اینترنت، بازار سهام را در معرض احساسات مشتری قرار داده، به دلیل نشر اخبار، بازار ثبات کمتری داشته و در معرض حملات مخرب است. بنابراین، تحقیقات بیشتر می‌تواند نقش مهمی در هموار نمودن مسیر تحلیل بازارهای سهام و افزایش قابلیت اطمینان آنها در آینده داشته باشد. یک مسیر تحقیقاتی امیدوار کننده، کشف الگوریتم‌های مختلف برای پیش‌بینی بلندمدت هست، زیرا بازارها در طولانی مدت مانند دستگاه‌های توزین عمل می‌کنند که نویز کمتری داشته و قابل پیش‌بینی هستند. با توجه به این نگاه، رویکردهای ترکیبی که تکنیک‌های آماری و یادگیری ماشینی را ترکیب می‌کنند، احتمالاً برای پیش‌بینی سهام مفیدتر هستند [۸].

هدف از انجام این پژوهش، یافتن روشی ترکیبی به منظور پیش‌بینی قیمت سهام است به گونه‌ای که چندین شرط برآورده شود: ۱) فرضیه‌های محدودکننده کمتری مورد نیاز باشد، ۲) قواعدی که استخراج می‌شود به راحتی قابل فهم و استفاده بوده و حتی‌الامکان تعداد آنها کم باشد و ۳) ضرورتی به حذف داده‌های جمع‌آوری شده اولیه نباشد. مدل پیشنهادی جهت تحقق اهداف ذکر شده، ترکیب روش تحلیل مؤلفه‌های اصلی[۲] با تئوری مجموعه‌های راف[۳] می‌باشد. در واقع دانش افزائی و نوآوری این پژوهش در ارائه مدلی ترکیبی با قدرت پیش‌بینی بالا و دیگر ویژگی‌های مثبت در حوزه پیش‌بینی قیمت سهام است.

با توجه به ضرورت مدل‌سازی برای پیش‌بینی تغییرات قیمت سهام، این مطالعه درصدد است به این سؤال پاسخ دهد که مدل ترکیبی تحلیل مؤلفه‌های اصلی و مجموعه‌های راف تا چه اندازه قادر به پیش‌بینی قیمت سهام بوده و در مقایسه با مدل‌های مشابه که از مجموعه‌های راف به تنهائی استفاده شده (به عنوان مثال پژوهش انجام شده قبلی توسط سارنج، کریمی و شهرامی بابکان (۱۳۹۶) [۹]) کدام‌یک از قدرت پیش‌بینی بیشتری برخوردار هستند؟ از این رو، هدف این پژوهش مقایسهٔ عملکرد این دو روش در پیش‌بینی قیمت سهام و دستیابی به مدلی مناسب برای پیش‌بینی قیمت روز آتی سهام در شرکت‌های پذیرفته شده در بورس اوراق بهادار

تهران است.

در ادامه، ابتدا مبانی نظری پژوهش شامل روش تحلیل مؤلفه‌های اصلی و تئوری مجموعه‌های راف ارائه شده و پیشینه پژوهش مرور می‌شود. سپس چارچوب مدل مورد استفاده ارائه می‌شود. بخش بعدی تجزیه‌وتحلیل داده‌ها است. در نهایت، پژوهش با نتیجه‌گیری و پیشنهادات به اتمام می‌رسد.

۲-مبانی نظری و پیشینه پژوهش

بازارهای مالی یکی از جذاب‌ترین ابداعات عصر ما هستند و تأثیر قابل توجهی برروی زمینه‌های مختلف نظیر کسب‌وکار، آموزش، مشاغل، تکنولوژی و در مجموع بر روی اقتصاد داشته‌اند [۱۰]. در طول سال‌ها، سرمایه‌گذاران و محققان به توسعه و آزمون مدل‌های رفتار قیمت سهام علاقه‌مند شده‌اند [۱۱]. به دلیل ماهیت پویا، غیرخطی، نامانایی، ناپارامتریک، نویز و آشفتگی، تحلیل بازار سهام و رفتار قیمت‌ها همراه با چالش‌های فراوان می‌باشد [۱۲]. بازار سهام تحت تأثیر فاکتورهای متعدد با وابستگی متقابل شامل متغیرهای اقتصادی، سیاسی، روانشناسی و خاص شرکت می‌باشد [۱۳]. تحلیل تکنیکی و بنیادی، دو رویکرد اصلی تحلیل بازارهای مالی هستند([١٤]؛ [١٥]) که مورد استفاده سرمایه‌گذاران به‌منظور تصمیم‌گیری در بازارهای مالی، سرمایه‌گذاری در سهام و دستیابی به سود بالا و ریسک پایین قرار می‌گیرد [١٦].

تحلیل بنیادی، عمدتاً مبتنی بر سه جنبه اساسی تحلیل اقتصاد کلان، تحلیل صنعت و تحلیل شرکت می‌باشد [۱۷]. رویکردهای ارزش‌گذاری متنوعی برای تحلیل بنیادی وجود دارد ([۱۸] ؛ [۱۹] ؛ [۲۰] ؛ [۲۱]).

برخی بر این باورند که قدیمی‌ترین شیوه نظام‌مند تحلیل تکنیکال متعلق به نظریه داو است که به سال ۱۸۸۴ باز می‌گردد[۲۲] . تحلیل تکنیکی بر پایه سه اصل اساسی استوار است که عبارتند از: لحاظ شدن همه چیز در قیمت، تمایل قیمت‌ها به حفظ روند فعلی به جای تغییر جهت و در نهایت اینکه تاریخ تکرار می‌شود. بیشتر پژوهش‌های انجام شده، تحلیل تکنیکال را کارآمد نشان می‌دهند، اما در پژوهش‌های گروهی دیگر، تحلیل تکنیکال کارآمد نبوده است ([۲۳] ؛ [٢٤]).

تحلیل و پیش‌بینی بازار سهام همچنان یک مسئله جالب و چالش برانگیز است. هرچه داده‌های بیشتری در دسترس قرار گیرد، ما در کسب و پردازش داده‌ها به منظور استخراج

دانش و بررسی تاثیرات بر عملکرد قیمت سهام با چالش‌های جدیدی روبه‌رو می‌شویم. این چالش‌ها شامل مواردی از قبیل معاملات الگوریتمی، پیش‌بینی‌های طولانی‌مدت و تجزیه‌وتحلیل احساسات می‌باشد. سیستم‌های معاملات الگوریتمی نحوه عملکرد بازارهای سهام را تغییر داده است. بیشتر حجم معاملات آتی سهام با استفاده از الگوریتم‌ها و نه توسط انسان تولید می‌شود. در حالی که معاملات الگوریتمی مزایایی مانند کاهش هزینه، کاهش تأخیر و عدم وابستگی به احساسات را به همراه دارد، اما چالش‌هایی را برای سرمایه‌گذاران خرد که فناوری لازم برای ساخت چنین سیستم‌هایی را ندارند، ایجاد می‌کند.

هدف از انجام این پژوهش، ارائه مدلی است که علاوه بر قابلیت اطمینان و سادگی استفاده و تفسیر، مستلزم استفاده از فرضیات غیرواقعی ساده کننده نبوده و بسته به نیاز کاربر، امکان ورود تعداد دلخواهی از متغیرها و در نتیجه سطح متفاوتی از پیچیدگی و دقت را داشته باشد. در ادامه، بخش‌های مرتبط با مبانی نظری تبیین می‌شوند. شایان ذکر است که با توجه به ویژگی‌های خاص پژوهش حاضر، این موارد هم با بخش مبانی نظری مرتبط هستند و هم به گونه‌ای با بخش روش پژوهش. با توجه به اینکه توضیحات بیشتر با جزئیات مرتبط در بخش روش پژوهش ارائه شده است، لذا برای استفاده هرچه بهتر خوانندگان، تبیین روش‌های مدل تحلیل مؤلفه‌های اصلی و مجموعه‌های راف در این بخش ارائه می‌شود.

۲-۱- روش تحلیل مؤلفه‌های اصلی

روش تحلیل مؤلفه‌های اصلی (که از این پس PCA نامیده می‌شود) تکنیکی برای کاهش داده‌ها یا کاهش بُعد متغیرها بوده که در رشته‌های مختلف نظیر پیش‌بینی سری‌های زمانی، تشخیص الگو، استخراج ویژگی، فشرده‌سازی داده‌ها و ... کاربردهای وسیعی داشته است. در زمینه کمّی مالی، PCA در کاوش سری‌های زمانی، استراتژی‌های معاملات دینامیک، محاسبات ریسک مالی، آربیتراژ آماری و پیش بینی قیمت سهام استفاده شده است [۲۵].

روش تحلیل مؤلفه‌های اصلی در سال ۱۹۰۱ برای اولین بار توسط کارل پیرسن معرفی شد. او توانست برای دو یا سه متغیر روش کاربردی ارائه دهد؛ در سال ۱۹۳۳ روش کاربردی برای محاسبه بیش از دو متغیر توسط هتلینگ ارائه شد. تحلیل مؤلفه‌های اصلی در تعریف ریاضی، یک تبدیل خطی متعامد است که داده‌ها را به دستگاه مختصات جدید می‌برد به طوری که بزرگترین واریانس داده بر روی اولین محور مختصات، دومین بزرگترین واریانس بر روی دومین محور مختصات قرار می‌گیرد و همین‌طور برای بقیه واریانس‌ها. تحلیل مؤلفه‌های اصلی

می‌تواند به‌منظور کاهش ابعاد داده‌ها مورد استفاده قرار گیرد؛ به این ترتیب، مؤلفه‌هایی از مجموع داده را که بیشترین تأثیر را در واریانس دارند، حفظ می‌کند. تحلیل مؤلفه‌های اصلی وقتی نتیجه بهتری خواهد داشت که متغیرهای اصلی تا حد زیادی همبستگی داشته باشند. اولین قدم در تحلیل مؤلفه‌های اصلی، محاسبه ماتریس کواریانس واریانس بین متغیرها است. با استفاده از مقادیر و بردارهای ویژه ماتریس کواریانس واریانس یا ماتریس همبستگی می‌توان مؤلفه‌های اصلی را تبیین نمود. هدف تحلیل مؤلفه‌های اصلی آن است که واریانس موجود در داده‌های چند متغیره را به مؤلفه‌هایی تبدیل نماید که مؤلفه‌های اصلی به ترتیب دارای واریانس از بزرگ به کوچک باشند [۲۶].

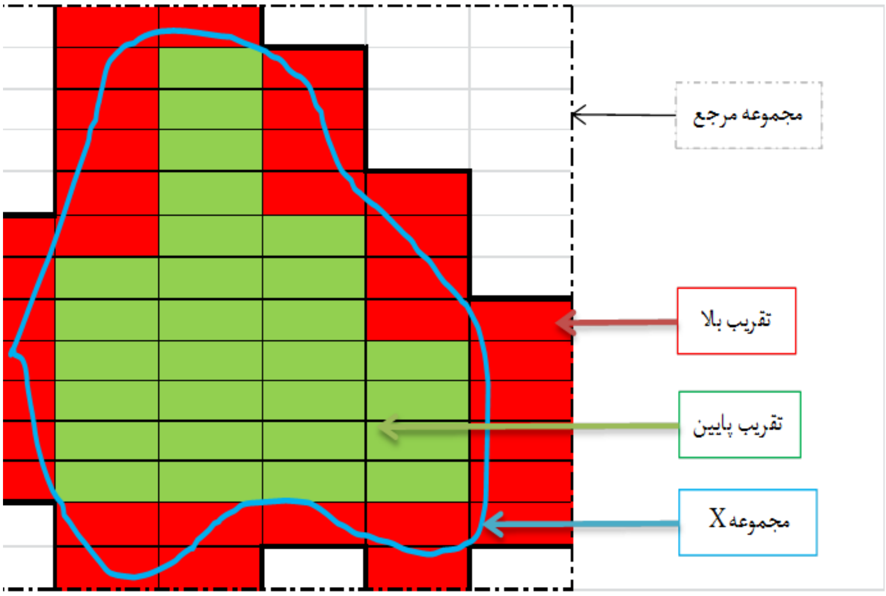
به عنوان مثالی ساده فرض کنیم با مجموعه‌ای از متغیرها (مثلاً ۲۰ متغیر) بررسی داده‌ها را شروع می‌کنیم. با اجرای این تکنیک در پایان به تعداد کمتری متغیر می‌رسیم که منعکس کننده بخش عمده اطلاعات مجموعه اولیه باشند. متغیرهای جدید، معمولاً ترکیب خطی از متغیرهای اولیه هستند. ایده اصلی در خصوص استفاده از PCA، کاهش ابعاد داده‌های مربوط به مجموعه‌ای از متغیرهای وابسته به‌گونه‌ای است که همچنان منعکس کننده تغییرات موجود در داده‌ها باشد. این هدف با تعریف مجموعه‌ای جدید از متغیرهای ناهمبسته (مؤلفه‌های اصلی) قابل دستیابی است. ردیابی مبدأ تکنیک‌های آماری، اغلب مشکل می‌باشد. اما به‌طور کلی شروع آنچه امروز تحت عنوان PCA شناخته می‌شود را به پیرسن و هتلینگ نسبت می‌دهند. فرآیند تبدیل متغیرها و کاهش بعد عمدتاً از سه طریق قابل انجام می‌باشد: روش اول به سادگی از طریق پیدانمودن ترکیبی خطی از متغیرها است که بیشترین تغییر در داده‌ها را توجیه می‌کند. روش دوم از طریق متغیرهای مکنون جدید (تحلیل عاملی اکتشافی) و روش سوم از طریق تحلیل عاملی تاییدی [۲۷].

واژه عامل در این مقوله با عامل در مباحثی نظیر تحلیل واریانس متفاوت بوده و معادل متغیر پنهان یا سازه می‌باشد. متغیر پنهان متغیری است که به طور مستقیم قابل اندازه‌گیری نبوده، بلکه غیرمستقیم از طریق متغیرهای قابل مشاهده اندازه‌گیری می‌شود. به‌عنوان مثال، اندازه‌گیری متغیر هوش (متغیر مکنون) با اندازه‌گیری برخی ویژگی‌ها (متغیرهای قابل اندازه‌گیری) نظیر حافظه کوتاه‌مدت و مهارت‌های کلامی، نوشتاری، خواندن، درک مطلب و ... انجام می‌گیرد. به‌منظور دستیابی به بهترین واریانس، داده‌ها بر روی یک زیرفضای ساخته شده توسط بردارهای ویژه تصویر می‌شود. در این صورت، مقدار ویژه متناظر با هر بردار ویژه، معادل میزان واریانس تبیین‌شده توسط آن بردار ویژه می‌باشد. از مهم‌ترین کاربردهای تحقیقاتی PCA می‌توان به رفع

هم‌خطی متغیرها و کاهش داده‌ها (تئوری سازی یا داده کاوی) اشاره نمود.

۲-۲- تئوری مجموعه‌های راف

تئوری مجموعه‌های راف در اوایل دهه ۱۹۸۰ توسط پاولاک معرفی شد. مجموعه‌های راف از ابزارهای مناسب برای کشف دانش در پایگاه‌های داده یا به عبارت دیگر ابزارهای ریاضی جهت کشف الگوهای مستتر در داده‌ها می‌باشد و به مرور زمان کاربردهای متنوعی نظیر محاسبات نرم، یادگیری ماشینی، تصمیم‌گیری، داده کاوی، سیستم‌های خبره و ... پیدا نمود. از آنجا که مجموعه‌های راف، اطلاعات اضافی در مورد داده‌ها به‌طور پیش فرض نیاز ندارند، مزایای زیادی نسبت به استدلال‌های آماری و احتمالاتی دارند. ورودی لازم به منظور تحلیل توسط مجموعه‌های راف به صورت جدول تصمیم می‌باشد. یک سیستم تصمیم، بیان کننده همه دانش در مورد مدل می‌باشد. جدول تصمیم ممکن است به دو دلیل به طور غیرضروری بزرگ باشد؛ یکی اینکه اعضای غیرقابل تمایز چند بار نمایش داده شده باشند و دیگری اینکه برخی مؤلفه‌ها ممکن است اضافی باشند. یک سیستم تصمیم ممکن است به دلیل وجود برخی مؤلفه‌های زائد در سیستم اطلاعاتی بی‌دلیل بزرگ باشد. تعریف و استفاده از دو مفهوم اساسی هسته و بی‌زائده، موجب حذف مؤلفه‌های غیرضروری و حفظ بخش حیاتی و ضروری سیستم اطلاعات می‌شود [۲۸]. در جدول تصمیم، هر ردیف نمایش دهنده یک وضعیت، رخداد یا نمونه می‌باشد و هر ستون یک مشخصه یا مؤلفه از نمونه‌ها را نمایش می‌دهد. مجموعه‌های راف تقریبی از مفاهیم غیرقطعی از طریق دو مجموعه تقریب بالا و تقریب پایین می‌باشد [۲۹] و قادر است قضاوت‌های ذهنی و مبهم متخصصان را حتی در حالتی که اندازه مجموعه داده‌ها کوچک و یا توزیع آن مشخص نباشد بررسی نماید [۳۰]. تقریب پایین شامل اعضایی می‌شود که قطعاً عضو مجموعه مورد نظر می‌باشند؛ اما، تقریب بالا در برگیرنده اعضایی است که احتمالاً عضو مجموعه می‌باشند. در شکل ۱ مفاهیم مرتبط با مجموعه‌های راف نمایش داده شده‌اند.



شکل ۱- نمایش مفاهیم مرتبط با مجموعه‌های راف

- مجموعه کلیه اعضایی که می‌تواند با اطمینان به عنوان اعضای X نسبت به R طبقه‌بندی شوند، تقریب پایین مجموعه X نسبت به R [٤] نامیده می‌شوند و با $R_*(X)$ نشان داده می‌شود، یعنی:

$$R * (X) = \{x: R(x) \subseteq X\}$$ رابطه ۱)

- مجموعه کلیه اعضایی که تنها می‌توان به عنوان اعضای محتمل X نسبت به R طبقه بندی نمود، تقریب بالای R [٥] مجموعه X نسبت به R نامیده می‌شود و با R(X) نشان داده می‌شود، یعنی:

$$R^*(X)= \{ x: R(x) \cap X \neq \emptyset\}$$ رابطه ۲)

- مجموعه کلیه اعضایی که نمی‌توان به طور قطع نه به عنوان عضو X و نه به عنوان عضو X نسبت به R طبقه بندی نمود، محدوده مرزی[۶] مجموعه X نسبت به R نامیده می‌شود و با $RN_R(X)$ نمایش داده می‌شود، یعنی:

$$RN_R(X) = R^*(X) - R_*(X).$$ رابطه ۳)

- یک مجموعه X نسبت به R راف[۷] نامیده می‌شود اگر و تنها اگر محدوده مرزی X تهی نباشد.

فرض کنیم T(U,A,C,D) یک جدول تصمیم می‌باشد که U مجموعه مرجع عناصر، A مجموعه از مشخصه‌های اصلی، C مجموعه‌ای از مؤلفه‌های شرطی، D مؤلفه تصمیم می‌باشد و C,D⊆A.

برای یک مجموعه دل‌خواه P⊆A، یک رابطه عدم تمایز به صورت زیر تعریف می‌شود:

رابطه ۴) $$IND(P) =\{ f(x; y) \in U\times U : \forall\ a \in P, a(x) = a(y)\}$$

رابطه عدم تمایز مجموعه مرجع را به زیرمجموعه‌های مجزا تقسیم می‌نماید. فرض کنیم U/IND(P) نشان دهنده خانواده‌ای از همه کلاس‌های هم ارزی تولید شده توسط IND(P) باشد. کلاس‌های هم ارزی U/IND(C) و U/IND(P) به ترتیب کلاس هم ارزی شرطی و تصمیم می‌باشند. در خصوص هر زیرمجموعه X⊆U ممکن است توصیف X توسط ترکیبی از کلاس‌های هم ارزی ممکن نباشد؛ یعنی نمایش صریح X با استفاده از اطلاعات در دسترس غیرممکن است. تئوری مجموعه‌های راف X را با استفاده تقریب‌های بالا و پایین توصیف می‌کند[۳۱].

یک ناحیه مثبت P از D، مجموعه‌ای از اعضای مجموعه مرجع U می‌باشد که با استفاده از مؤلفه‌های P با اطمینان می‌توان به یکی از طبقات U/IND(P) نسبت داد،

رابطه ۵) $$POS_P(D)=\bigcup_{x\in U/IND(D)} \underline{P}X$$

وابستگی مولفه‌ها (مثلاً وابستگی D در P به صورت زیر تعریف می‌شود:

رابطه ۶) $$\gamma_{p}(D)=\frac{|POSp(D)|}{|U|}$$

که $|A|$ عدد اصلی مجموعه A می‌باشد.

مشخصه a ∈ C در C غیرضروری است اگر

رابطه ۷) $$\gamma_{p}(D)=\gamma_{p-a}(D)$$

در غیر این‌صورت a یک مشخصه ضروری در P نسبت به D می‌باشد. یک مجموعه دلخواه B⊆C مستقل نامیده می‌شود اگر همه مشخصه‌های آن ضروری باشد.

با استفاده از تعاریف بالا، یک مجموعه بی‌زائده از مشخصه‌ها را می‌توان به این صورت تعریف نمود: یک مجموعه از مشخصه‌های $R \subseteq C$ بی زائده نامیده می‌شود اگر R مستقل بوده و $POS_R(D)=POS_C(D)$. بی‌زائده، مجموعه‌ای از مشخصه‌ها است که همان تقسیم‌بندی ایجاد شده توسط C را حفظ می‌کند. با این تفاسیر، یک بی‌زائده، کوچک‌ترین زیرمجموعه از مشخصه‌ها است که طبقه بندی مشابه مجموعه مرجع از عناصر تولید می‌کند. به عبارت دیگر، مشخصه‌ای که به بی‌زائده تعلق نداشته باشد، از نظر طبقه‌بندی عناصر در مجموعه، اضافی می‌باشد. روش‌های متعددی جهت تولید بی‌زائده‌های مجموعه مرجع نظیر الگوریتم‌های هیوریستیک وجود دارد [۳۲].

۲-۳- پیشینه تجربی

با بررسی پیشینه پژوهش اعم از کشور ایران و سایر کشورها مشخص می‌شود که مطالعاتی که به‌طور کامل و دقیقاً مرتبط با موضوع پژوهش حاضر باشند، خیلی کم هستند. اما در مجموع، تلاش شد تا در بخش پیشینه پژوهش مرتبط‌ترین مطالعات ارائه شوند. در ادامه برخی از مطالعات مرتبط با حوزه پیش‌بینی قیمت سهام، مجموعه روش‌های مرتبط با این پیش‌بینی، مجموعه‌های راف، برخی مدل‌های ترکیبی و ... ارائه می‌شوند.

روسو[۸] (۲۰۰۳) [۳۳] تعدادی از روش‌های مورد استفاده در پیش‌بینی قیمت سهام – شامل روش‌های کلاسیک که توسط تحلیل‌گران بنیادی و تکنیکی استفاده شده تا روش‌های جدیدتر- را تحلیل نمود. تفاوت‌های اساسی بین روش‌های پیش‌بینی بر اساس میزان پیچیدگی، محدودیت‌ها، الزامات و دقت هر روش وجود دارد. هر روش مناسب شرایط خاصی می‌باشد و انتخاب روش بهینه که متناسب با آن شرایط خاص بوده و به طور کامل داده‌های موجود را ارزیابی نماید بسیار اهمیت دارد. در این مطالعه، شرایط عملی که هر روش کارایی دارد، معرفی شده است.

القاهری و همکاران (۲۰۰۸)[۳٤] مدلی برای پیش‌بینی قیمت سهام براساس مجموعه‌های راف ارائه نمودند. به‌منظور افزایش کارایی فرایند پیش‌بینی، مجموعه‌های راف با الگوریتم گسسته‌سازی استدلالی بولین مورد استفاده قرار گرفتند. مجموعه‌های راف قابلیتِ یافتن حداقل زیرمجموعه از مؤلفه‌های حاوی حداکثر اطلاعات (بی‌زائده‌ها) و معرفی قواعد تصمیم را دارند. قواعد تصمیم استخراج شده در مورد پیش‌بینی حرکات قیمتی یکی از سهم‌های موجود در بازار بورس کویت اعمال شد. مقایسه نتایج حاصل از کاربرد مدل مورد استفاده با الگوریتم

شبکه‌های عصبی نشان داد که روش مجموعه‌های راف ضمن استفاده از تعداد قواعد کمتر، دارای دقت بالاتری نیز می‌باشد.

چن و همکاران[۹] (۲۰۱۰) [۳۵] با تلفیق مجموعه‌های راف، رویکرد توزیع احتمال تجمعی[۱۰]، رویکرد اصل حداقل آنتروپی[۱۱] و الگوریتم ژنتیک[۱۲]، مدلی ترکیبی را به منظور استخراج قواعد پیش‌بینی بازده سهام در بازار بورس تایوان ارائه نمودند. نتایج تجربی حاکی از برتری مدل از نظر دقت و بازده بالاتر نسبت به روش‌های مقایسه‌ای (خرید و نگهداری، RST و GA) می‌باشد.

سای و سایو[۱۳] (۲۰۱۰) [۳۶] به منظور تشخیص و فیلترنمودن متغیرهای اضافی در تصمیم‌گیری از سه روش تحلیل مؤلفه‌های اصلی، الگوریتم ژنتیک و درخت تصمیم CART[۱۴] استفاده نمودند. شبکه‌های عصبی انتشار بازگشتی به‌عنوان ابزار پیش‌بینی مدل استفاده شد. این مدل‌ها قادر بودند۸۰ درصد از متغیرهای اولیه اضافی را فیلتر نمایند. متغیرهای جدید به مثابه عامل‌های مهم برای پیش‌بینی قیمت سهام می‌باشند که در تصمیمات سرمایه‌گذاری آتی مورد استفاده قرار می‌گیرند.

اسپریج[۱۵] (۲۰۱۱) [۳۷] با تلفیق PCA و Fuzzy-SVR[۱۶] مدلی برای پیش‌بینی قیمت سهام معرفی نمود. سهم‌های با روند تاریخی مشابه، با استفاده از تحلیل مؤلفه‌های اصلی انتخاب شدند. برای تعریف تابع چگالی قیمت سهام، اطلاعات لازم به صورت فازی تعریف شد. به منظور ایجاد یک تابع رگرسیون پیش‌بینی قیمت سهام، از رگرسیون بردار پشتیبان استفاده شد. این روش برای نمونه‌های با اندازه‌های مختلف و با هر توزیع اختلال مناسب بوده و فرآیندهای اصلاحی پیچیده سایر روش‌های شبکه‌های عصبی را حذف می‌کند. به علاوه استفاده از اطلاعات فازی، شکل خروجی‌های پیش‌بینی را از یک نقطه به یک بازه با احتمال متناظر تغییر می‌دهد.

فلاح‌پور، گل‌ارزی و فتوره‌چیان (۱۳۹۲) [۳۸] در پژوهشی یک مدل ترکیبی ماشین بردار پشتیبان بر پایه الگوریتم ژنتیک برای پیش‌بینی روند حرکتی قیمت سهام ارائه کردند. نتایج پژوهش آن‌ها نشان داد که این مدل ترکیبی در مقایسه با روش ماشین بردار پشتیبان ساده، از عملکرد و دقت بالاتری از برخوردار است.

شاکری و مراد پور (۱۳۹۴) [۳۹] مدلی را به‌منظور پیش‌بینی قیمت سهام بورس اوراق بهادار تهران با استفاده از نظریه مجموعه‌های راف ارائه نمودند. در این مدل تحرکات بازار سهام مورد تحلیل و بررسی قرار گرفته تا سرمایه‌گذاران بتوانند با استفاده از این مدل برای خرید یا فروش سهام خود تصمیم درست اتخاذ نمایند. مدل پیشنهادی بر پایه تئوری راف بنا

شده و پس از گسسته‌سازی داده‌ها، از قابلیت مجموعه‌های راف برای یافتن ویژگی‌هایی که اهمیت بیشتری در طبقه‌بندی داده‌ها و تولید قواعد تصمیم داشته و در نهایت استخراج قواعد تصمیم استفاده نموده است. در این مطالعه با استفاده از داده‌های هفت سال بازار بورس تهران (۱۳۹۰ - ۱۳۸۴) نتیجه‌گیری شده که مجموعه‌های راف می‌تواند ابزار اثربخشی برای رسیدن به هدف پیش‌بینی صحیح با مقدار قابل قبولی از سطح اطمینان باشد.

جای وانگ و جون وانگ[۱۷] (۲۰۱۴) [٤٠] شبکه‌های عصبی تابع احتمال مؤثر زمانی و تحلیل مؤلفه‌های اصلی را ترکیب نموده و مدل حاصل را به‌منظور پیش‌بینی قیمت سهام به کار بردند. در بخش یادگیری مدل، ابتدا از روش PCA برای استخراج مؤلفه‌های اصلی داده‌های ورودی و سپس از[۱۸] STNN در پیش‌بینی سری زمانی قیمت‌ها استفاده شد. مقایسه نتایج حاصل از مدل پیشنهادی با سایر مدل‌ها نظیر شبکه‌های عصبی پس انتشار[۱۹] (BPNN)، PCA-BPNN و STNN، نشان داد که مدل پیشنهادی عملکرد بهتری دارد. به علاوه، قدرت پیش‌بینی مدل در خصوص شاخص‌های SSE، HS۳۰۰، S&P۵۰۰ و DJIA مورد ارزیابی قرار گرفت.

در مطالعات دیگری نیز مدل‌هایی با ترکیب روش‌های PCA و شبکه‌های عصبی با هدف پیش‌بینی رفتار قیمت سهام پیشنهاد شده‌اند ([٤١] ؛ [٤٢] ؛ [٤٣]).

روش تحلیل مؤلفه‌های اصلی در مدیریت پورتفلیو نیز مورد مطالعه قرار گرفته است. یانگ (۲۰۱۵)[۲۰] [٤٤] کاربرد PCA را در بازار سهام استرالیا با استفاده از شاخص ASX۲۰۰ و اجزای آن از آوریل ۲۰۰۰ تا فوریه ۲۰۱۴ مطالعه نموده است. مؤلفه‌های اصلی اول و دوم به ترتیب نمایش دهنده منابع ریسک و بالاترین همبستگی با شاخص ASX۲۰۰ می‌باشد. واریانس تبیین شده توسط مؤلفه اصلی اول به عنوان معیاری از سطح ریسک سیستماتیک استفاده شد و می‌تواند علامت هشدار بحران مالی باشد.

افسر و هلیل (۱۳۹۶) [٤٥] با هدف بهینه‌سازی سبد سهام، مدلی را با رویکرد ترکیبی روش‌های تحلیل تکنیکال و داده‌کاوی ارائه نمودند؛ در این مدل، پتانسیل آتی سهام با در نظر گرفتن شاخص‌های تحلیل تکنیکال به وسیله شبکه عصبی فازی پیش‌بینی شد و بر اساس پیش‌بینی‌های به دست آمده، مدل ریاضی بهینه‌سازی برمبنای عواملی چون میانگین، واریانس و چولگی سبد سهام ارائه و در نهایت با استفاده از الگوریتم ژنتیک حل شد. در پژوهشی با هدف مشابه، محمدی و همکاران (۱۳۹۶) [٤٦] به‌منظور شناسایی متغیرهای مؤثر و کارا در فرآیند تشکیل سبد سهام بهینه، با تلفیق روش‌های تحلیل پوششی داده‌ها و الگوریتم جستجوی ارگانیسم‌های هم‌زیست، مدلی ارائه نمودند.

سارنج، کریمی و شهرامی بابکان (۱۳۹۶) [۹] با استفاده از ترکیب روش‌های مختلف گسسته‌سازی داده‌ها و روش‌های متنوع تولید بی‌زائده و مجموعه‌های راف، قواعدی را به منظور پیش‌بینی قیمت سهام بانک صادرات ایران در بازه شش سال متوالی به طور مجزا استخراج نمودند. نتایج اعتبارسنجی مدل در سال‌های مختلف با روندهای متفاوت داده‌ها، نشان داد که ترکیب خاصی از روش‌های گسسته‌سازی و تولید بی‌زائده، قابلیت پیش‌بینی قیمت سهام را با دقت بالا دارند. مقایسه بازده حاصل از اعمال این مدل نسبت به روش خرید و نگهداری نیز برتری کامل آن را نشان می‌دهد.

فخاری و همکاران (۱۳۹۶) [٤٧] با هدف مقایسهٔ عملکرد مدل‌های پیش‌بینی شبکهٔ عصبی با مدل‌های کلاسیک و معرفی مدل مناسب برای پیش‌بینی قیمت روز آتی سهام، از داده‌های قیمت روزانهٔ بازار و شاخص‌های تکنیکی مالی به عنوان متغیرهای ورودی استفاده نمودند؛ همچنین، برای طراحی مدل آریما، داده‌های قیمت بسته‌شدن روزانه به عنوان متغیر ورودی و قیمت بسته‌شدن روز آتی به عنوان متغیر خروجی هر دو مدل در دورهٔ زمانی ۱۳۹۰ تا ۱۳۹۳ در نظر گرفته شد. نتایج به دست آمده از شبکهٔ عصبی بیزین بیان کنندهٔ خطای کمتر و قدرت پیش‌بینی بیشتر آن در مقایسه با مدل آریما است. یافته‌های تحقیق گویای کارایی بیشتر شبکهٔ عصبی بیزین در استفاده از فرصت‌های سرمایه‌گذاری کوتاه‌مدت بازار است که می‌تواند به سرمایه‌گذاران در انتخاب پرتفوی مناسب و کسب بازده بیشتر کمک کند.

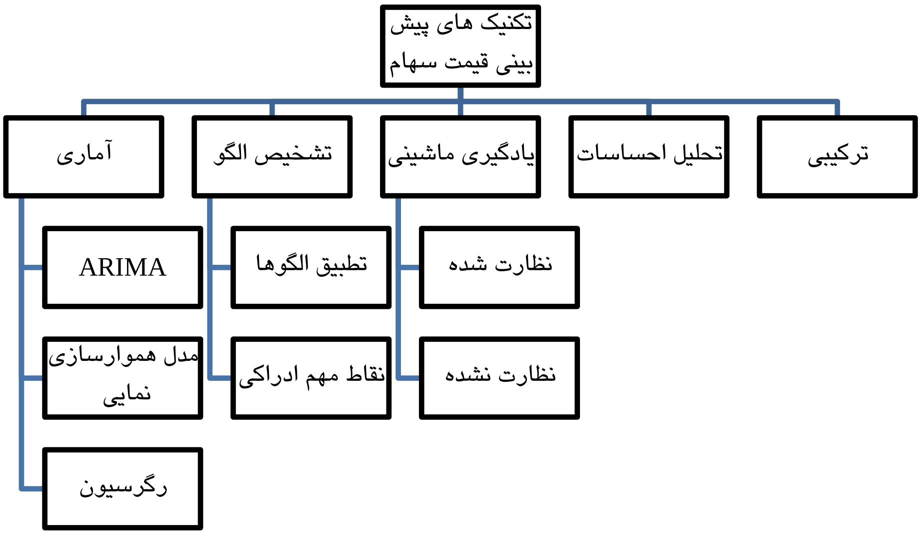
پورزمانی و رضوانی اقدم (۱۳۹۶) [٢٤] در پژوهشی با استفاده از استراتژی‌های تکنیکال میانگین متحرک نمایی (EMA) و شاخص قدرت اندازه نسبی حرکت (RSI) و مقایسه آن با روش خرید و نگهداری، به خرید سهام پرداختند. نتایج پژوهش آنها نشان داد که در دوره‌های به شدت صعودی استراتژی‌های تکنیکال کارایی لازم را ندارند، اما در دوره‌ها با روند متعادل، استراتژی‌های تکنیکال برای خرید سهام کاراتر است.

افشاری راد، علوی و سینایی (۱۳۹۷) [٤٨] در پژوهشی روند سهام را با روشهای تحلیل تکنیکال و روش‌های هوشمند یادگیری ماشین روی شاخص کل پیش‌بینی کردند. به این ترتیب که ابتدا داده‌های مورد نیاز جمع‌آوری شده، سپس به ۲۵ روش تحلیل داده می‌شوند، از میان این ۲۵ روش، ده روش با اولویت طبق روش انتخاب ویژگی کاهش ابعاد، انتخاب می‌شوند؛ خروجی این مرحله به پنج روش هوشمند یادگیری ماشین، ماشین بردار پشتیبان خطی و کرنل گوسی، درخت تصمیم، نزدیک‌ترین K همسایه و بیز ساده داده می‌شود. سپس، برای تصمیمگیری نهایی از روش رأی اکثریت استفاده شده است. یافته‌ها حاکی از این است که نرخ

پیش‌بینی صحیح روش پیشنهادی به طور متوسط ۹۷ درصد بوده است.

غلامیان و داودی (۱۳۹۷) [٤٩] با استفاده از رویکرد جنگل تصادفی، از روش‌های طبقه‌بندی هوش مصنوعی به همراه شاخص‌های تکنیکی شامل شاخص قدرت نسبی قیمت، استوکاستیک، حجم تعادل موازنه شده، ویلیامز، بازدهی روزانه و شاخص مکدی (MACD) برای پیش‌بینی روند قیمت در بازار سهام استفاده کرده‌اند. نتیجه پژوهش روی داده‌های روزانه شاخص بورس اوراق بهادار تهران در سالهای ۱۳۹۳ تا ۱۳۹۵ نشان داد که دقت روش پیشنهادی در برآورد روند بازار ۶۴ درصد است.

شاه و همکاران[۲۱] (۲۰۱۹) [۸]، ابتدا یک بررسی اجمالی از بازارهای سهام و طبقه‌بندی روش‌های پیش‌بینی بازار سهام ارائه داده و سپس بر برخی از دستاوردهای تحقیق در تجزیه‌وتحلیل سهام و پیش‌بینی تمرکز نموده‌اند. در ادامه رویکردهای تکنیکی، بنیادی، کوتاه‌مدت و بلندمدت مورد استفاده برای تجزیه‌وتحلیل سهام و در نهایت چالش‌ها و فرصت‌های تحقیقاتی در این زمینه را تشریح نموده‌اند (شکل۲).


تکنیک های پیش بینی قیمت سهام
ترکیبی
تحلیل احساسات
یادگیری ماشینی
تشخیص الگو
آماری
نظارت شده
نظارت نشده
تطبیق الگوها
نقاط مهم ادراکی
ARIMA
مدل هموارسازی نمایی
رگرسیون

شکل ۲ طبقه‌بندی تکنیک‌های پیش‌بینی قیمت سهام (منبع: [۲۲])

پیمانی و همکاران (۱۳۹۹) [۲۲] در پژوهشی به بررسی سودآوری نمودارهای شمعی در شرایط مختلف پرداخته‌اند. بدین منظور، با استفاده از داده‌های مربوط به قیمت‌های روزانه سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران طی ۱۵سال از آبان ۱۳۸۲ تا آبان ۱۳۹۷، میزان بازدهی و درصد موفقیت سرمایه‌گذاری بر اساس ۱۳ نمودار شمعی مختلف در دو افق زمانی یک‌روزه و ۱۰روزه و در شرایط مختلف صعودی و نزولی و حالت‌های مختلف گردش معامله‌ها اندازه‌گیری شده است. یافته‌های پژوهش به این ترتیب ذکر شده‌اند: بدون مدنظر قرار دادن روند یا گردش معامله‌ها، شمعِ فروش در دوره نگهداری یک‌روزه و شمع‌های خرید در دوره نگهداری ۱۰روزه، بیشترین بازدهی را (بیش از هزینه‌های معامله‌ها) به‌همراه دارند. با بررسی نمودارها در روندهای صعودی و نزولی، عملکرد نمودارها افزایش خواهد یافت. همچنین با در نظر گرفتن گردش معامله‌ها، بهترین بازدهی، مربوط به شمعِ خرید در روند صعودی و در گروه شرکت‌های با ارزش معامله‌های متوسط و دوره نگهداری ۱۰روزه است. از سوی دیگر، نتایج پژوهش دلیلی بر توصیه به اتخاذ دیدگاه کوتاه‌مدت در دوره‌های نزولی است.

سیف و همکاران (۱۴۰۰) [۲۳] با بررسی داده‌های شاخص کل بازار سهام ایران از تاریخ ۱۳۸۷/۰۲/۲۵ تا ۱۳۹۹/۰۹/۰۵ به‌طور روزانه و استفاده از نوسان‌نمای موج الیوت نشان دادند که در شاخص بورس اوراق بهادار تهران، شناسایی امواج الیوت امکان‌پذیر است و الگوریتم‌های ماشین بردار پشتیبان و درخت تصمیم، قابلیت پیش‌بینی روند شاخص کل را برای آینده با دقت بالای ۹۰ درصد دارند.

مقایسه اهم پیشینه پژوهش به‌طور خلاصه در جدول ۱ ارائه شده است. همان‌طور که قبلاً ذکر شد، هدف از انجام این پژوهش، ارائه مدلی است که علاوه بر قابلیت اطمینان و سادگی استفاده و تفسیر، مستلزم استفاده از فرضیات غیرواقعی ساده کننده نبوده و بسته به نیاز کاربر، امکان ورود تعداد دلخواهی از متغیرها و در نتیجه سطح متفاوتی از پیچیدگی و دقت را داشته باشد. این کار برای نخستین بار صورت می‌پذیرد.

جدول ۱ – خلاصه پیشینه پژوهش

منبع	هدف	متغیرها و تکنیک‌های مورد استفاده	یافته‌ها
[۹]	پیش‌بینی قیمت سهام	ترکیب روش‌های مختلف گسسته‌سازی داده‌ها و روش‌های متنوع تولید بی‌زائده – مجموعه‌های راف	قابلیت پیش‌بینی قیمت سهام با دقت بالا، بازده بالاتر مدل نسبت به روش خرید و نگهداری
[۲۳]	پیش‌بینی روند شاخص کل بورس اوراق بهادار تهران	نوسان‌نمای موج الیوت، الگوریتم‌های ماشین بردار پشتیبان و درخت تصمیم	قابلیت پیش‌بینی با دقت بالای ۹۰٪

[۲٤]	پیش‌بینی قیمت سهام	استراتژی‌های تکنیکال میانگین متحرک نمایی (EMA) و شاخص قدرت اندازه نسبی حرکت(RSI)	درمقایسه با روش خرید و نگهداری، در دوره‌های به شدت صعودی، استراتژی‌های تکنیکال کارایی لازم را ندارند، اما در دوره‌ها با روند متعادل، استراتژی‌های تکنیکال برای خرید سهام کاراتر است.
[۳٤]	پیش‌بینی قیمت سهام	مجموعه‌های راف، الگوریتم گسسته‌سازی استدلالی بولین	روش مجموعه‌های راف در مقایسه با الگوریتم شبکه‌های عصبی، ضمن استفاده از تعداد قواعد کمتر، دارای دقت بالاتری می‌باشد
[۳۵]	استخراج قواعد پیش‌بینی بازده سهام	تلفیق مجموعه‌های راف، رویکرد توزیع احتمال تجمعی[۱۰]، رویکرد اصل حداقل آنتروپی[۱۱] و الگوریتم ژنتیک[۱۲]	برتری مدل از نظر دقت و بازده بالاتر نسبت به روش‌های مقایسه‌ای (خرید و نگهداری، RST و GA)
[۳٦]	تشخیص و فیلترنمودن متغیرهای اضافی در تصمیم‌گیری (پیش‌بینی قیمت سهام)	تحلیل مؤلفه‌های اصلی، الگوریتم ژنتیک، درخت تصمیم CART[۱۴]، شبکه‌های عصبی انتشار بازگشتی.	۸۰ درصد از متغیرهای اولیه اضافی فیلتر شدند. متغیرهای جدید به مثابه عامل‌های مهم برای پیش‌بینی قیمت سهام می‌باشند
[۳۷]	پیش‌بینی قیمت سهام	تحلیل مؤلفه‌های اصلی، تابع چگالی قیمت سهام به صورت فازی، رگرسیون بردار پشتیبان	این روش برای نمونه‌های با اندازه‌های مختلف و با هر توزیع اختلال، مناسب بوده و فرآیندهای اصلاحی پیچیده سایر روش‌های شبکه‌های عصبی را حذف می‌کند. بعلاوه استفاده از اطلاعات فازی، شکل خروجی‌های پیش‌بینی را از یک نقطه به یک بازه با احتمال متناظر تغییر می‌دهد
[۳۸]	پیش‌بینی روند حرکتی قیمت سهام	مدل ترکیبی ماشین بردار پشتیبان بر پایه الگوریتم ژنتیک	مدل ترکیبی در مقایسه با روش ماشین بردار پشتیبان ساده، از عملکرد و دقت بالاتری از برخوردار است
[۳۹]	پیش‌بینی قیمت سهام بورس اوراق بهادار تهران	مجموعه‌های راف	مجموعه‌های راف می‌تواند ابزار اثربخشی برای رسیدن به هدف پیش‌بینی صحیح با مقدار قابل قبولی از سطح اطمینان باشد
[٤۰]	پیش‌بینی قیمت سهام	ترکیب شبکه‌های عصبی تابع احتمال مؤثر زمانی (STNN) و تحلیل مؤلفه‌های اصلی	مدل پیشنهادی نسبت به با سایر مدل‌ها نظیر شبکه‌های عصبی پس انتشار[۱۹] (BPNN)، PCA-BPNN و STNN، عملکرد بهتری دارد.

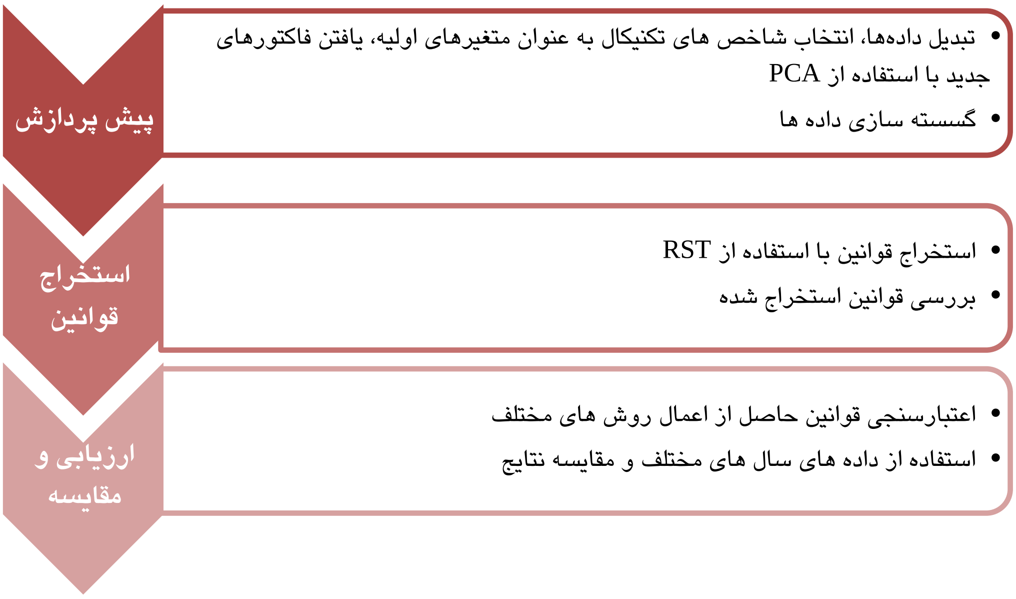
منبع	هدف	متغیرها و تکنیک‌های مورد استفاده	یافته‌ها
[٤٤]	مدیریت پرتفلیو	PCA، شاخص ASX۲۰۰ در بازار سهام استرالیا	مؤلفه‌های اصلی اول و دوم به ترتیب نمایش دهنده منابع ریسک و بالاترین همبستگی با شاخص ASX۲۰۰ می‌باشد. واریانس تبیین شده توسط مؤلفه اصلی اول به عنوان معیاری از سطح ریسک سیستماتیک استفاده شد و می‌تواند علامت هشدار بحران مالی باشد.
[٤٧]	مقایسهٔ عملکرد مدل‌های پیش‌بینی شبکهٔ عصبی با مدل‌های کلاسیک و معرفی مدل مناسب برای پیش‌بینی قیمت روز آتی سهام	داده‌های قیمت روزانهٔ بازار و شاخص‌های تکنیکال، شبکهٔ عصبی بیزین، مدل آریما، قیمت بسته‌شدن روزانه، قیمت بسته‌شدن روز آتی	نتایج به دست آمده از شبکهٔ عصبی بیزین بیان کنندهٔ خطای کمتر و قدرت پیش‌بینی بیشتر آن در مقایسه با مدل آریما است
[٤٨]	پیش‌بینی روند قیمت سهام براساس شاخص کل	روش تحلیل تکنیکال-روش‌های هوشمند یادگیری ماشینی، ماشین بردار پشتیبان خطی و کرنل گوسی، درخت تصمیم، روش رأی اکثریت	یافته‌ها حاکی از این است که نرخ پیش‌بینی صحیح روش پیشنهادی به طور متوسط ۹۷ درصد بوده است
[٤٩]	پیش‌بینی روند قیمت در بازار سهام	رویکرد جنگل تصادفی- روش‌های‌طبقه‌بندی‌هوش مصنوعی به همراه شاخص‌های تکنیکی شامل شاخص قدرت نسبی قیمت، استوکاستیک، حجم تعادل موازنه شده، ویلیامز، بازدهی روزانه و شاخص	نتیجه پژوهش روی داده‌های روزانه شاخص بورس اوراق بهادار تهران در سالهای ۱۳۹۳ تا ۱۳۹۵ نشان داد که دقت روش پیشنهادی در برآورد روند بازار ۶۴ درصد است.

۳- روش‌شناسی پژوهش

در بخش‌های پیشین، توضیحات مبسوطی در خصوص روش تحلیل مؤلفه‌های اصلی و تئوری مجموعه‌های راف ارائه شد. در ادامه، سایر موارد مرتبط با روش پژوهش تشریح می‌شوند. ابتدا مدل پیشنهادی ارائه و نحوه استفاده از آن شامل پیش‌پردازش، استخراج قوانین و ارزیابی و مقایسه توضیح داده می‌شوند. پژوهش حاضر از نظر هدف کاربردی و از نظر روش توصیفی - تحلیلی می‌باشد.

۳-۱- مدل پیشنهادی

چارچوب مدل مورد استفاده پژوهش حاضر در شکل ۳ نمایش داده شده است.



پیش پردازش

- تبدیل داده‌ها، انتخاب شاخص های تکنیکال به عنوان متغیرهای اولیه، یافتن فاکتورهای جدید با استفاده از PCA
- گسسته سازی داده ها

استخراج قوانین

- استخراج قوانین با استفاده از RST
- بررسی قوانین استخراج شده

ارزیابی و مقایسه

- اعتبارسنجی قوانین حاصل از اعمال روش های مختلف
- استفاده از داده های سال های مختلف و مقایسه نتایج

شکل ۳ -چارچوب مدل پیشنهادی

۳-۲- فرایند اجرای مدل

نحوه استفاده از چارچوب مدل مفهومی مطابق با اجرای مراحل ذیل می‌باشد:

۳-۲-۱- تبدیل داده‌ها، انتخاب شاخص‌های تکنیکال اساسی به عنوان متغیرهای اولیه، یافتن متغیرهای جدید با استفاده از PCA

در گام نخست، داده‌های اولیه از سایت شرکت مدیریت فناوری بورس تهران (www.tsetmc.com) استخراج شد و محاسبات شاخص‌های تکنیکال مورد نظر به عنوان متغیرهای اولیه بر اساس آن انجام گرفت. تا این مرحله، فرآیند انجام کار مشابه پژوهش پیشین انجام گرفته توسط سارنج و همکاران (۱۳۹۶)[۹] می‌باشد؛ لکن، به منظور کاهش متغیرها و در نتیجه استخراج قواعد تصمیم‌گیری ساده‌تر، با استفاده از تحلیل مؤلفه‌های اصلی و تحلیل عاملی، متغیرهای جدیدی با تعداد کمتر تولید می‌شود که ضمن حفظ خواص داده‌های اولیه، تصمیم گیری را سهل‌تر نماید. شایان ذکر است که متغیرهای ایجاد شده به عنوان مؤلفه‌های شرطی[۲۲] و نوسان قیمت سهم در روز آتی به عنوان مؤلفه تصمیم[۲۳] انتخاب شدند. به طور

خلاصه، خروجی این مرحله "جدول تصمیم[۲۴]" می‌باشد.

۳-۲-۲- گسسته‌سازی داده‌ها

با توجه به اینکه داده‌های اولیه به صورت سری زمانی و با ماهیت پیوسته می‌باشند، در این مرحله با استفاده از تکنیک‌های موجود، داده‌ها به بازه‌های گسسته تبدیل می‌شوند. این مرحله از پردازش بر روی عامل‌های ایجاد شده حاصل از تحلیل مؤلفه‌های اصلی به عنوان مؤلفه‌های شرطی انجام می‌گیرد. در خصوص مؤلفه تصمیم، داده‌ها به سه متغیر زبانی U (افزایش قیمت)، D (کاهش قیمت) و N (بدون تغییر) نسبت به روز تصمیم‌گیری (پیش‌بینی حالات احتمالی مختلف تغییر قیمت سهام در روز بعد به منظور تصمیم‌گیری در خصوص خرید، فروش یا نگهداری سهام) گسسته می‌شوند.

۳-۲-۳- استخراج قوانین با استفاده از تئوری مجموعه‌های راف

از جدول تصمیم با فرمت خاص به عنوان ورودی نرم افزار Rosetta به منظور تولید بی‌زائده‌ها[۲۵] و استخراج قواعد استفاده می‌شود.

۳-۲-۴- ارزیابی و مقایسه قواعد استخراج شده حاصل از اعمال روش‌های مختلف

داده‌های مورد استفاده به دو قسمت (گروه یادگیری و گروه کنترل) تقسیم شده و قواعد استخراج شده با استفاده از داده‌های یادگیری در مورد داده‌های کنترل استفاده می‌شوند. بررسی دقتِ قواعد و میزان تأیید و پشتیبانی آنها توسط داده‌های موجود در این مرحله انجام می‌گیرد. علاوه بر آن، قواعد حاصل از ترکیب روش‌های مختلف گسسته‌سازی و تولید بی‌زائده با هم مقایسه می‌شوند.

۳-۲-۵- استفاده از داده‌های سال‌های مختلف و مقایسه نتایج

مدل بر روی داده‌های با دوره زمانی مختلف مربوط به شرکت ایران خودرو، اجرا و نتایج با هم مقایسه می‌شود. در مرحله اول، بازه انتخابی داده‌ها، یک سال شمسی (فاصله زمانی بین برگزاری دو مجمع عمومی عادی سالیانه متوالی) یعنی از ۱۳۹۷/۰۵/۰۳ لغایت ۱۳۹۸/۰۴/۲۷ می‌باشد و نتایج اعمال هر دو مدل (مدل RST و مدل ترکیبی RST-PCA) بر روی داده‌ها بررسی می‌شود. در مرحله بعد، به‌منظور ارزیابی پایایی مدل، بازه انتخابی داده‌ها به سال ۱۳۸۳

لغایت ۱۳۹۸ گسترش داده شده و نتایج مقایسه می‌شود. شایان ذکر است در مطالعه قبلی انجام گرفته توسط سارنج و همکاران (۱۳۹۶) [۹] ، مدل RST بر روی داده‌های بانک صادرات اعمال و نتایج آن در مجله راهبرد مدیریت مالی منتشر شد. در این تحقیق، مدل ترکیبی پیشنهادی بر روی داده‌های مذکور اعمال و نتایج حاصل شده از دو مدل با هم مقایسه می‌شود.

۴- تحلیل داده‌ها و یافته‌های پژوهش

داده‌های مربوط به سهم ایران خودرو از سال ۱۳۸۳ تا ابتدای سال ۱۳۹۸ از سایت شرکت مدیریت فناوری بورس تهران استخراج و فرآیند پردازش داده‌ها مطابق مراحل توضیح داده شده در بالا انجام گرفت. به این ترتیب که با استفاده از داده‌های اولیه تعدادی از شاخص‌های تکنیکی (میانگین متحرک (MA)، مومنتم (MTM)، %K تصادفی[۲۶]، شاخص قدرت نسبی (RSI) و %R ویلیامز) برای کل دوره محاسبه شد. در یک مرحله، مدل را بر روی داده‌های حد فاصل برگزاری دو مجمع عمومی سالیانه شرکت اعمال نموده و در مرحله دیگر به منظور بررسی نقش طول دوره زمانی، مدل بر روی کل داده‌ها اعمال شد. زمان برگزاری مجمع عمومی سی و یکم تیرماه هر سال می‌باشد که به همین دلیل از داده‌های بازه زمانی ۱۳۹۷/۰۵/۰۳ لغایت ۱۳۹۸/۰۴/۲۷ استفاده شد. جدول شماره ۲ آماره‌های توصیفی داده‌های بازه زمانی یک ساله داده‌ها را نشان می‌دهد.

جدول ۲ – آماره‌های توصیفی داده‌های بازه ۱۳۹۸–۱۳۹۷

Analysis N	Std. Deviation	Mean	
۲۱۳	۲۶.۰۶۸۱۴	۳۸.۵۰۹۳	k
۲۱۳	۲۳۱.۷۷۶۷	۲۹۵۷.۹۳۲	ma
۲۱۳	۶.۱۹۳۲۵	۹۸.۷۴۹	mtm
۲۱۳	۲۶.۰۶۸۱۴	-۶۱.۴۹۰۷	r
۲۱۳	۱۴.۶۹۹۳۷	۴۴.۹۰۱۳	rsi

جدول شماره ۳ ماتریس همبستگی متغیرهای اولیه (شاخص‌های تکنیکال) را نشان می‌دهد.

جدول ۳ – ماتریس همبستگی متغیرها

		K	ma	mtm	r	rsi
Correlation	k	۱	-۰.۱۳	۰.۷۸	۱	۰.۷۶۲
	ma	-۰.۱۳	۱	-۰.۰۲۵	-۰.۱۳	۰.۰۱۹
	mtm	۰.۷۸	-۰.۰۲۵	۱	۰.۷۸	۰.۹۰۴
	r	۱	-۰.۱۳	۰.۷۸	۱	۰.۷۶۲
	rsi	۰.۷۶۲	۰.۰۱۹	۰.۹۰۴	۰.۷۶۲	۱

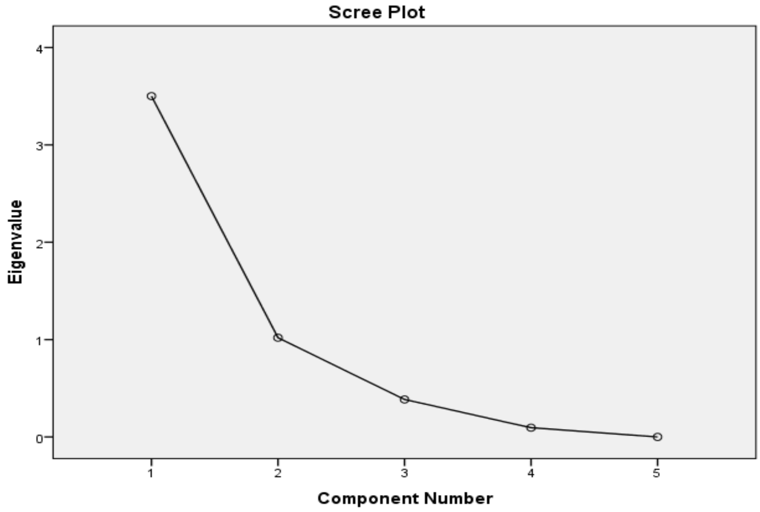
نتایج اعمال روش تحلیل مؤلفه‌های اصلی در جدول ۴ و شکل‌های ۴ و ۵ آمده است. آزمون اسکری (شکل ۳) نشان می‌دهد که دو عامل با مقادیر ویژه نزدیک به عدد یک و بالاتر می‌تواند برای داخل شدن به تحلیل مورد بررسی قرار گیرد. همان‌طور که در جدول ۵ و شکل‌های ۴ و ۵ مشخص است، توانسته‌ایم تعداد متغیرها را از ۵ به ۲ عامل (متغیر مکنون) کاهش دهیم ولی با این مزیت که قادر به تبیین بیش از ۹۰٪ واریانس داده‌های اولیه باشد. توجه به شکل ۴ و جدول ۴ حاکی از آن است که در مختصات دوران یافته متعامد، دو عامل جدید با استفاده از ترکیب خطی متغیرهای اولیه قابل تشخیص می‌باشند که در ادامه با رویکرد تحلیل عاملی اکتشافی به نامگذاری آن‌ها می‌پردازیم.

جدول ۴ – واریانس تبیین شده توسط عامل‌های جدید مربوط به داده‌های ۱۳۹۸–۱۳۹۷

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
۱	۳.۵۰۱	۷۰.۰۲۱	۷۰.۰۲۱	۳.۵۰۱	۷۰.۰۲۱	۷۰.۰۲۱	۳.۴۸۹	۶۹.۷۸۳	۶۹.۷۸۳
۲	۱.۰۱۹	۲۰.۳۸۴	۹۰.۴۰۶	۱.۰۱۹	۲۰.۳۸۴	۹۰.۴۰۶	۱.۰۳۱	۲۰.۶۲۳	۹۰.۴۰۶
۳	۰.۳۸۵	۷.۶۹۷	۹۸.۱۰۳						
۴	۰.۰۹۵	۱.۸۹۷	۱۰۰						
۵	۱.۰۰e-۱۳	۱.۰۲e-۱۳	۱۰۰						

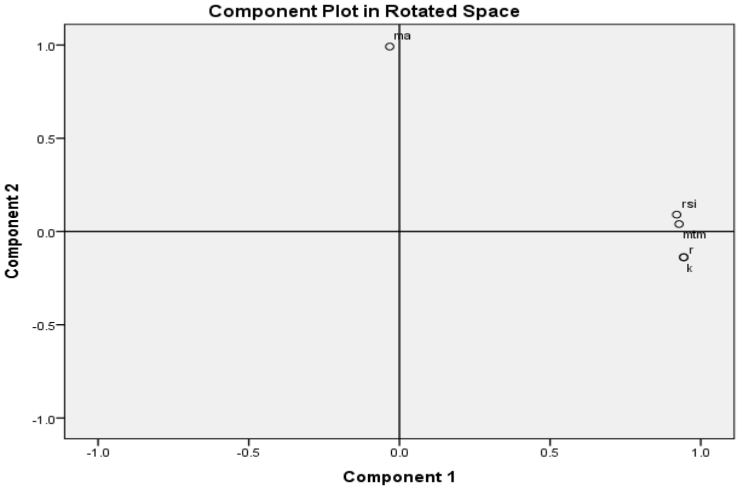
Extraction Method: Principal Component Analysis.

آزمون اسکری در شکل ۴ نشان می‌دهد که دو عامل با مقادیر ویژه نزدیک به عدد یک و بالاتر می‌تواند برای داخل شدن به تحلیل مورد بررسی قرار گیرد.



شکل ۴ – نمودارِ اسکری و تعیین تعداد عامل‌های جدید

همان‌طور که در شکل ۵ مشخص است، تعداد پنج متغیر اولیه به ۲ عامل جدید (متغیر مکنون) به گونه‌ای کاهش یافته که قادر به تبیین بیش از ۹۰٪ واریانس داده‌های اولیه باشد.



شکل ۵ – نمایش مؤلفه‌های اصلی بر اساس متغیرهای اولیه

با استفاده از ضرایب داده شده در جدول ۵ به عنوان ضرایب تبدیل متغیرهای اولیه به عامل‌ها، داده‌های مربوط به عامل‌های جدید تولید و در محاسبات بعدی مورد استفاده قرار می‌گیرد.

جدول ۵ – ماتریس ضرایب اجزا در عامل‌های جدید

	Component	
	۱	۲
k	۰.۲۶۶	−۰.۰۸۹
ma	۰.۰۳۸	۰.۹۶۸
mtm	۰.۲۷	۰.۰۸۴
r	۰.۲۶۶	−۰.۰۸۹
rsi	۰.۲۷	۰.۱۳۳

مراحل گسسته‌سازی و خوشه‌بندی داده‌ها، تقسیم داده‌ها به دو گروه یادگیری و کنترل، تولید بی‌زائده‌ها و استخراج قواعد تصمیم، یک مرتبه برروی عامل‌های جدید (که ترکیبی از متغیرهای اولیه می‌باشند) و مرتبه بعد به طور مستقیم بر روی متغیرهای اولیه (شاخص‌های تکنیکال) اجرا شد که نتایج آن به ترتیب در جدول‌های ۶ و ۷ نمایش داده شده است. همان‌طور که قبلاً ذکر شد، قواعد تصمیم براساس داده‌های یادگیری مدل استخراج و سپس بر روی داده‌های کنترل راستی آزمایی می‌شود. در جداول ۶، ۷ و۸، قدرت پیش‌بینی مدل بر روی داده‌های کنترل نشان داده شده است؛ به‌عنوان مثال، در جدول ۶، از مجموع ۳۴ داده کنترل واقعی مربوط به کاهش سهام (D) در روز آتی، مدل قادر بوده ۳۳ مورد (۹۷ درصد) را به درستی تشخیص داده و در خصوص یک مورد پیش‌بینی خاصی مبنی بر کاهش یا افزایش سهم ارائه ننموده است. همچنین، از مجموع ۳۰ داده کنترل واقعی مربوط به افزایش سهام (U) در روز آتی، مدل توانسته ۲۶ مورد (۸۷ درصد) را به درستی تشخیص داده و در خصوص چهار مورد پیش بینی خاصی ارائه ننموده است. در مجموع، ۹۲ درصد پیش‌بینی‌ها صحیح بوده و نکته قابل تاکید این است که ۸ درصد باقیمانده، عدم پیش بینی بوده و در مجموع، مدل پیش بینی اشتباه نداشته است.

جدول ۶ – نتایج اعتبارسنجی مدل ترکیبی PCA و RST بر روی داده‌های بازه ۱۳۹۸-۱۳۹۷

	Predicted				
Actual		D	U	Undefined	
	D	۳۳	۰	۱	۰.۹۷
	U	۰	۲۶	۴	۰.۸۷
	Undefined	۰	۰	۰	Undefined
		۱.۰	۱.۰	۰.۰	۰.۹۲

جدول ۷ – نتایج اعتبارسنجی مدل RST بر روی داده‌های بازه ۱۳۹۸-۱۳۹۷

	Predicted				
Actual		D	U	Undefined	
	D	۳۰	۰	۴	۰.۸۸
	U	۰	۲۵	۵	۰.۸۳
	Undefined	۰	۰	۰	Undefined
		۱.۰	۱.۰	۰.۰	۰.۸۶

به‌منظور انجام مقایسه‌های تکمیلی، مدل برروی داده‌های بازه زمانی ۱۳۸۳ لغایت ۱۳۹۸ اجرا گردید که نتایج نهایی مطابق جدول ۸ می‌باشد. جدول شماره ۹ آماره‌های توصیفی داده‌های بازه زمانی پانزده ساله داده‌ها را نشان می‌دهد. در این مورد نیز مدل ترکیبی نتایج مطلوب‌تری نسبت به مدل RST ارائه نمود و عامل‌های جدید (مومنتم و میانگین متحرک) مشابه حالت داده‌های سالیانه می‌باشد. بررسی جداول ۸ و ۱۰ نشان دهنده نوسان شدیدتر داده‌ها و قدرت تبیین کمتر عامل‌ها در مقایسه با داده‌های بازه یک ساله می‌باشد. به بیان ساده‌تر، قدرت پیش‌بینی مدل با افزایش بازه زمانی داده‌های یادگیری از یک سال به چند سال، نسبت عکس دارد که این موضوع با کاربرد اصلی تحلیل تکنیکال منافاتی نداشته و نوسانات شدیدتر داده‌ها در بازه‌های زمانی طولانی‌تر می‌تواند یکی از علل اصلی باشد.

جدول ۸ – نتایج اعتبارسنجی مدل ترکیبی PCA و RST بر روی داده‌های بازه ۱۳۹۸–۱۳۸۳

	Predicted					
Actual		U	D	N	Undefined	
	U	۳۱۴	۱	۰	۱۸۸	۰.۶۲
	D	۱	۳۰۵	۰	۱۷۸	۰.۶۳
	N	۱	۰	۵	۵۷	۰.۰۸
	Undefined	۰	۰	۰	۰	Undefined
		۰.۹۹	۰.۹۹	۱.۰	۰.۰	۰,۵۹

جدول ۹ – آماره‌های توصیفی داده‌های بازه ۱۳۹۸–۱۳۸۳

	Mean	Std. Deviation	Analysis N
k	۵۰.۰۰۵۴	۴۰.۰۷۶۷۴	۳۵۰۲
ma	۲۹۳۱.۱۸	۱۴۷۷.۴۹۷	۳۵۰۲
mtm	۱۰۰.۶۰۲۸	۱۲.۲۹۶۷۲	۳۵۰۲
r	-۴۹.۹۶۶	۴۰.۰۷۶۷۲	۳۵۰۲
rsi	۴۹.۴۳۵۷	۲۸.۳۵۶۲۹	۳۵۰۲

جدول ۱۰ – واریانس تبیین شده توسط عامل‌های جدید مربوط به داده‌های ۱۳۹۸–۱۳۸۳

Component	**Initial Eigenvalues**			**Extraction Sums of Squared Loadings**			**Rotation Sums of Squared Loadings**		
	Total	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
۱	۲.۸۲۷	۵۶.۵۳۹	۵۶.۵۳۹	۲.۸۲۷	۵۶.۵۳۹	۵۶.۵۳۹	۲.۸۲۱	۵۶.۴۱۶	۵۶.۴۱۶
۲	۱.۰۱	۲۰.۲۰۶	۷۶.۷۴۵	۱.۰۱	۲۰.۲۰۶	۷۶.۷۴۵	۱.۰۱۶	۲۰.۳۲۹	۷۶.۷۴۵
۳	۰.۸۴	۱۶.۸۰۳	۹۳.۵۴۹						
۴	۰.۳۲۲	۶.۴۳۴	۹۹.۹۸۳						
۵	۰.۰۰۱	۰.۰۱۷	۱۰۰						

خلاصه‌ای از نتایج تحلیل‌های انجام شده بر روی داده‌های بازه ۱۳۹۸–۱۳۹۷ در جدول شماره ۱۱ ارائه شده است. کاهش تعداد قواعد پیش‌بینی استخراج شده از ۴۷۱ به ۱۸۸و افزایش درصد پیش‌بینی‌های صحیح از ۸۶٪ به ۹۲٪ از مزایای استفاده از مدل ترکیبی PCA-RST نسبت به مدل RST می‌باشد.

جدول ۱۱ – خلاصه نتایج اعتبارسنجی روش‌های مختلف تحلیل داده‌های بازه ۱۳۹۸-۱۳۹۷

ردیف	روش گسسته‌سازی داده‌ها	روش تولید بی‌زائده	نتایج مربوط به مدل ترکیبی PCA و RST		نتایج مربوط به مدل RST	
			درصد پیش‌بینی‌های صحیح	تعداد قواعد استخراج شده	درصد پیش‌بینی‌های صحیح	تعداد قواعد استخراج شده
۱	Naïve Algorithm…	Genetic Algorithm…	۹۲	۱۸۸	۸۶	۴۷۱

۵- نتیجه‌گیری و پیشنهادات

در این پژوهش چندین هدف پیگیری می‌شد که نتایج حاصل شده تاحد زیادی حاکی از نیل به آن اهداف می‌باشد. به بیان ساده، مجموع این اهداف، به دنبال یافتن روشی به منظور پیش‌بینی قیمت سهام، بدون پیش فرض‌های الزام آور و محدود کننده، با حفظ داده‌های اولیه، دقت بالا، سهولت استفاده و تفسیر و حتی‌الامکان تعداد کمتری از قواعد تصمیم‌گیری می‌باشد. استفاده از روش تئوری مجموعه‌های راف (RST)، نیاز به انجام فرضیات محدودکننده را رفع نموده و ترکیب آن با روش تحلیل مؤلفه‌های اصلی (PCA) باعث حفظ داده‌های اولیه در قالب عامل‌های جدید (که ترکیب خطی از متغیرهای اولیه مطابق جدول ۴ می‌باشند) شده است. استفاده از مدل پیشنهادی هم باعث بهبود دقت قواعد تصمیم و هم موجب کاهش تعداد آنها و سهولت استفاده شده که جدول شماره ۶ موید این موضوع می‌باشد. مزیت اساسی که در این پژوهش تحلیل عاملی اکتشافی ایجاد نموده است، قابلیت نامگذاری متغیرهای جدید است؛ با بررسی ضرایب متغیرهای اولیه در عامل‌های جدید در جدول شماره ۴ این موضوع مهم قابل نتیجه‌گیری است که پنج متغیر اولیه را می‌توان، با حفظ بخش عمده خواص داده‌های اولیه، به دو عامل جدید کاهش داد (مطابق جدول شماره ۴ بیش از ۹۰٪ واریانس توسط این دو عامل تبیین می‌شود)؛ عامل اول را می‌توان عامل مومنتم نامگذاری نمود و عامل دوم تقریباً معادل متغیر اولیه میانگین متحرک می‌باشد که این موضوع نقش مهمی در کاهش تعداد قواعد تصمیم و ملموس بودن استفاده از آنها دارد.

به‌منظور بررسی پایایی مدل پیشنهادی، دو سری داده دیگر به عنوان ورودی مدل استفاده شدند. یک سری داده مربوط به بازه زمانی سال‌های ۱۳۹۸-۱۳۸۳ شرکت ایران‌خودرو بود که

مطابق جدول ۸ قدرت پیش‌بینی قواعد استخراج شده تقریباً ۶۰درصد است؛ در این حالت نیز، اگرچه نسبت به مدل RST عملکرد مطلوبتری دارد، لکن نسبت به داده‌های بازه زمانی یک ساله دارای قدرت پیش‌بینی کمتری هستند. مقایسه آماره‌های توصیفی نشان دهنده نوسانات شدیدتر در بازه زمانی طولانی‌تر بوده به گونه‌ای که قدرت تبیین عامل مومنتم حدود ۲۰ درصد کاهش یافته است. سری داده دوم مربوط به بانک صادرات می‌باشد که موضوع پژوهش قبلی سارنج و همکاران (۱۳۹۶) [۹] بود. در این مورد هم استفاده از مدل ترکیبی منجر به ایجاد عامل‌های مشابه (مومنتم و میانگین متحرک) و همچنین نتایج مطلوبتری از نظر قدرت پیش‌بینی و تعداد کمتر قواعد شده است. به منظور انجام تحقیقات آتی مرتبط با این پژوهش، پیشنهاد می‌گردد این مدل با تعداد بیشتری از شاخص‌های تکنیکال و با کمک برنامه‌های کامپیوتری در معاملات الگوریتمی استفاده شود. به‌علاوه مقایسه مجموعه‌های راف با سایر روش‌های موجود نیز مفید می‌باشد.

٦- پی‌نوشت‌ها

١. Machine Learning
٢. Principal Component Analysis (PCA)
٣. Rough Set Theory (RST)
٤. R-lower approximation
٥. R-upper approximation
٦. Boundary region
٧. Rough (inexact)
٨. Rusu
٩. Cheng, C. H., Chen, T. L., & Wei, L. Y.
١٠. Cumulative probability distribution approach (CDPA)
١١. Minimize the entropy principle approach (MEPA)
١٢. Genetic algorithms (GAs)
١٣. Tsai, C. F., & Hsiao, Y. C.
١٤. Classification and Regression Tree
١٥. Ma, X., & Spreij, P. J. C.
١٦. Fuzzy support vector regression
١٧. Wang, J., & Wang, J.
١٨. Stochastic time effective neural network
١٩. Backpropagation neural network
٢٠. Yang, L.
٢١. Shah, D., Isah, H., & Zulkernine, F.
٢٢. Conditional attributes
٢٣. Decision attribute
٢٤. Decision table
٢٥. Reducts
٢٦. Stochastic %K (%K)

۷-منابع

[١] Bernstein, S., Lebow, R. N., Stein, J. G., & Weber, S. (٢٠٠٠). God gave physics the easy problems: Adapting social science to an unpredictable world. European

Journal of International Relations, ٦(١), ٧٦-٤٣.

[٢] amiri, M., aghaei, M. (٢٠٢١). Using Set Covering Approach for Decision-Making Criteria Classification while Correlation between Criteria Exist. *Management Research in Iran*, ٢١(٣), ١-٢٣.

[٣] Malkiel, B. G. (٢٠٠٣). The efficient market hypothesis and its critics. Journal of economic perspectives, ١٧(١), ٨٢-٥٩.

[٤] Lo, A. W., & MacKinlay, A. C. (٢٠١١). A non-random walk down Wall Street. Princeton University Press.

[٥] Basu, S. (١٩٨٣). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. Journal of financial economics, ١٢(١), ١٥٦-١٢٩.

[٦] Fama, E. F., & French, K. R. (١٩٨٨). Permanent and temporary components of stock prices. Journal of political Economy, ٩٦(٢), ٢٧٣-٢٤٦.

[٧] Lakonishok, J., Shleifer, A., & Vishny, R. W. (١٩٩٤). Contrarian investment, extrapolation, and risk. The journal of finance, ٤٩(٥), ١٥٧٨-١٥٤١.

[٨] Shah, D., Isah, H., & Zulkernine, F. (٢٠١٩). Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, ٧(٢), ٢٦.

[٩] saranj, A., karimi, T., Shahrami Babakan, M. (٢٠١٧). The Application of Rough Set Theory in Stock Price Forecasting (Case Study: Iran Saderat Bank). Financial Management Strategy, ٥(٣), ١٤٤-١١٩. doi: ١٠,٢٢٠٥١/jfm.٢٠١٧,١٢٦٨٠,١١٨٩ (in Persian).

[١٠] Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (٢٠١٨). NSE stock market prediction using deep-learning models. Procedia computer science, ١٣٢, ١٣٦٢-١٣٥١.

[١١] Fama, E. F. (١٩٩٥). Random walks in stock market prices. Financial analysts journal, ٥١(١), ٨٠-٧٥.

[١٢] Abu-Mostafa, Y. S., & Atiya, A. F. (١٩٩٦). Introduction to financial forecasting. Applied intelligence, ٦(٣), ٢١٣-٢٠٥.

[١٣] Zhong, X., & Enke, D. (٢٠١٧). Forecasting daily stock market return using dimensionality reduction. Expert Systems with Applications, ٦٧, ١٣٩-١٢٦.

[١٤] Park, C. H., & Irwin, S. H. (٢٠٠٧). What do we know about the profitability of technical analysis?. Journal of Economic surveys, ٢١(٤), ٨٢٦-٧٨٦.

[١٥] Nguyen, T. H., Shirai, K., & Velcin, J. (٢٠١٥). Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications, ٤٢(٢٤), -٩٦٠٣

٩٦١١.

[١٦] Arévalo, R., García, J., Guijarro, F., & Peris, A. (٢٠١٧). A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting. Expert Systems with Applications, ٨١, ١٩٢-١٧٧.

[١٧] Hu, Y., Liu, K., Zhang, X., Su, L., Ngai, E. W. T., & Liu, M. (٢٠١٥). Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review. Applied Soft Computing, ٣٦, ٥٥١-٥٣٤.

[١٨] Gordon, M. J., & Shapiro, E. (١٩٥٦). Capital equipment analysis: the required rate of profit. Management science, ٣(١), ١١٠-١٠٢.

[١٩] Gordon, M. J. (١٩٥٩). Dividends, earnings, and stock prices. The review of economics and statistics, ١٠٥-٩٩.

[٢٠] Imam, S., Barker, R., & Clubb, C. (٢٠٠٨). The use of valuation models by UK investment analysts. European Accounting Review, ١٧(٣), ٥٣٥-٥٠٣.

[٢١] Dutta, A., Bandopadhyay, G., & Sengupta, S. (٢٠١٢). Prediction of stock performance in the Indian stock market using logistic regression. International Journal of Business and Information, ٧(١), ١٠٥.

[٢٢] Peymany Foroushany, Moslem; Erza, Amir Hossein; Salehi, Mahdi; Salehi, Ahmad (١٣٩٩). Trades Return Based on Candlestick Charts in Tehran Stock Exchange. Financial Research, (٢٢) ١, ٨٩-٦٩.

[٢٣] Seif, Samira; Jamshidi Navid, Babak; Ghanbari, Mehrdad; Ismailpour, Mansour. (١٤٠٠). Predicting the stock market trend in Iran using Elliott wave oscillation and relative strength index. Financial research. (١) ٢٣. ١٥٧-١٣٤.

[٢٤] Pourzamani, Zahra; Rezvani Aghdam, Mohsen (١٣٩٦). Comparison of the effectiveness of combined strategies of technical analysis with the method of buying and holding for buying stocks in the uptrend and downtrends. Quarterly Journal of Financial Research in Securities Analysis, (١٠) ٣٣, ٣١-١٧.

[٢٥] Ghorbani, M., & Chong, E. K. (٢٠٢٠). Stock price prediction using principal components. PloS one, ١٥(٣), e٠٢٣٠١٢٤.

[٢٦] Azar, Adel; Khadivar, Amene (١٣٩٨). Application of multivariate statistical analysis in management, Negah Danesh, Tehran, third edition.

[٢٧] Beaumont, R. (٢٠١٢). An introduction to principal component analysis & factor analysis using SPSS ١٩ and R (psych package). Factor Analysis and Principal Component Analysis (PCA), ٢٤(٩-٨).

[٢٨] Sadeghi Moghadam, M., Karimi, T., bandesi, S. (٢٠٢١). Service Supply Chain Risk Assessment Applying Rough Set Theory Approach: Case of Payment Service

Providers. *Management Research in Iran*, ٢٢(١), ٦٩-٩٤.

[٢٩] Pawlak, Z. (٢٠١٢). Rough sets: Theoretical aspects of reasoning about data (Vol. ٩). Springer Science & Business Media.

[٣٠] Shahraki, A., Tahmasbi Abdar, Z. (٢٠١٩). Failure Mode and Effects Analysis Using Rough Set Theory and Grey Relational Projection Method. *Modern Research in Decision Making*, ٤(٢), ١-٣٥.

[٣١] Fazayeli, F., Wang, L., & Mandziuk, J. (٢٠٠٨, October). Feature selection based on the rough set theory and expectation-maximization clustering algorithm. In International Conference on Rough Sets and Current Trends in Computing (pp. ٢٨٢-٢٧٢). Springer, Berlin, Heidelberg.

[٣٢] Chouchoulas, A., & Shen, Q. (٢٠٠١). Rough set-aided keyword reduction for text categorization. Applied Artificial Intelligence, ١٥(٩), ٨٧٣-٨٤٣.

[٣٣] Rusu, V., & Rusu, C. (٢٠٠٣). Forecasting methods and stock market analysis. Creative Math, ١٢, ١١٠-١٠٣.

[٣٤] Al-Qaheri, H., Hassanien, A. E., & Abraham, A. (٢٠٠٨). Discovering stock price prediction rules using rough sets. Neural Network World, ١٨(٣), ١٨١.

[٣٥] Cheng, C. H., Chen, T. L., & Wei, L. Y. (٢٠١٠). A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. Information Sciences, ١٨٠(٩), ١٦٢٩-١٦١٠.

[٣٦] Tsai, C. F., & Hsiao, Y. C. (٢٠١٠). Combining multiple feature selection methods for stock prediction:union, intersection, and multi-intersection approaches. Decision Support Systems, ٥٠(١), ٢٦٩-٢٥٨.

[٣٧] Ma, X., & Spreij, P. J. C. (٢٠١١). PCA-Fuzzy-SVR Stock Price Prediction.

[٣٨] Fallahpour, Saeed; Gol Arzi, Gholam Hossein; Fatoreh Chian, Nasser (١٣٩٢). Predicting the moving trend of stock prices using a support vector machine based on genetic algorithm in Tehran Stock Exchange. Financial Research, ١٥ (٢), ٢٦٩ - ٢٨٨.

[٣٩] Shakeri, Mehdi; Moradpour, Mona. (١٣٩٤). Tehran Stock Exchange Stock Price Forecast Using Rough Set Theory, ٣rd International Conference on Accounting and Management, Tehran, https://civilica.com/doc/٤٤١٧٢٣.

[٤٠] Wang, J., & Wang, J. (٢٠١٥). Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks. Neurocomputing, ١٥٦, ٧٨-٦٨.

[٤١] Ince, H., & Trafalis, T. B. (٢٠٠٧). Kernel principal component analysis and support vector machines for stock prediction. IIE Transactions, ٣٩(٦), ٦٣٧-٦٢٩.

[٤٢] Guo, Z., Wang, H., Yang, J., & Miller, D. J. (٢٠١٥). A stock market forecasting model combining two-directional two-dimensional principal component analysis and radial basis function neural network. PloS one, ١٠(٤), e٠١٢٢٣٨٥.

[٤٣] Zahedi, J., & Rounaghi, M. M. (٢٠١٥). Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran Stock Exchange. Physica A: Statistical Mechanics and its Applications, ٤٣٨, ١٨٧-١٧٨.

[٤٤] Yang, L. (٢٠١٥). An application of principal component analysis to stock portfolio management.

[٤٥] Afsar, A., Helyel, F. (٢٠١٧). A Hybrid Approach to Portfolio Optimization Using Technical Analysis and Data Mining. *Modern Research in Decision Making*, ٢(٢), ١-٢٢.

[٤٦] Mohammadi, S., Mohammadi, E., Barzinpour, F. (٢٠١٨). Portfolio Optimization in Tehran Stock Exchange by Using Data Envelopment Analysis and Symbiotic Organisms Search. *Modern Research in Decision Making*, ٣(٢), ٢٢٣-٢٤٨.

[٤٧] Fakhari, Hussein; Valipour Khatir, Mohammad; Mousavi, Seyedeh Maedeh. (١٣٩٦). Investigating Performance of Bayesian and Levenberg-Marquardt Neural Network in Comparison Classical Models in Stock Price Forecasting. Financial Research, (٢) ١٩. ٣١٨-٢٩٩.

[٤٨] Afshari Rad, Elham; Alawi, Sayyid Enayatullah; Sinayi, Hassan Ali (١٣٩٧). An intelligent model for predicting stock prices using technical analysis methods. Financial Research, (٢٠) ٢, ٢٦٤-٢٤٩.

[٤٩] Golamian, Elham; Davoodi, Mohammad Reza (١٣٩٧). Predicting the price trend in the stock market using a random forest algorithm. Journal of Financial Engineering and Securities Management, ٩ (٣٥), ٣٢٢-٣٠١.