فصلنامــه علمی پژوهشی دانش سرمایه‌گــذاری
سال دهم/ شماره سی‌وهفتم/ بهار ۱۴۰۰

طراحی موتور استنتاج سیستم خبره جهت ارزیابی و انتخاب صندوق سرمایه‌گذاری در ایران مبتنی بر ویژگی‌های صندوق: رویکرد تئوری راف

رضا تهرانی
استاد گروه مدیریت مالی، دانشکده مدیریت، دانشگاه تهران.
rezatehrani@ut.ac.ir

محمدعلی میرزا کوچک شیرازی
دانشجوی دکتری مدیریت مالی، دانشکده مدیریت، دانشگاه تهران، پردیس بین المللی کیش (نویسنده مسئول)
mohammadalishirazi@yahoo.com

سید مجتبی میرلوحی
استادیار دانشکده صنایع و مدیریت، دانشگاه صنعتی شاهرود.
Mirlohism@shahroodut.ac.ir

تاریخ دریافت: ۹۸/۰۱/۲۷ تاریخ پذیرش: ۹۸/۰۲/۰۲

چکیده

با توجه به تعداد زیاد صندوق‌های سرمایه‌گذاری، عملکرد متفاوت آنها در گذشته و ویژگی‌های متفاوت آنها، انتخاب صندوق سرمایه‌گذاری مناسب توسط یا برای سرمایه‌گذاران امری مهم و حیاتی است. علاوه بر عملکرد گذشته شامل ریسک و بازده، عوامل متعدد ساختاری و رفتاری دیگری وجود دارد که می‌تواند بر بازدهی آتی صندوق سرمایه‌گذاری اثر گذارد. این عوامل می‌تواند شامل اندازه، هزینه‌های صندوق، حق الزحمه مدیریت، نسبت گردش پرتفولیوی، جریان نقدینگی، عمر صندوق، میزان تمرکز، موقعیت مکانی، تجربه و مهارت مدیریت، جنسیت مدیریت، حجم معاملات، ساختار حاکمیتی صندوق‌های سرمایه‌گذاری و ... باشد. از سوی دیگر، در صورت اطلاع از عوامل اثرگذار بر تصمیم‌گیری، فرآیند تحلیل، تفسیر و انتخاب صندوق نیز موضوعی ناشناخته است که براساس خبرگی فرد تصمیم‌گیرنده صورت می‌پذیرد. بنابراین مسأله اصلی این تحقیق علاوه بر تعیین ویژگی‌های مؤثر در عملکرد آینده صندوق‌های سرمایه‌گذاری در ایران، طراحی یک موتور استنتاج جهت تعیین وضعیت صندوق‌های سرمایه‌گذاری در آینده و ایجاد یک سیستم خبره تصمیم‌گیری در این حوزه می‌باشد. بر اساس نتایج به دست آمده از تئوری راف، مشخصه‌های موقعیتی و اثرگذار بر عملکرد صندوق‌های سرمایه‌گذاری تعیین گردیده و مدلهای قوانین برای تصمیم گیری در خصوص انتخاب صندوق سرمایه گذاری استخراج شد. بر اساس نتایج به دست آمده بی زائده های هر الگو استخراج گردید و الگوهایی که بالاترین درصد اعتبار را داشتند

معرفی گردیدند. بر اساس نتایج به دست آمده یک سیستم خبره طراحی شد که به عنوان پشتیبان تصمیم در انتخاب صندوق‌های سرمایه گذاری می تواندکمک‌کننده باشد.

واژه‌های کلیدی: صندوق سرمایه‌گذاری، مشخصه های صندوق‌های سرمایه‌گذاری، سیستم خبره، تئوری راف، شاخص شارپ، شاخص جنسن، شاخص ترینر.

۱- مقدمه

موفقیت هر فعالیت سرمایه‌گذاری به دانش و توان سرمایه‌گذاران در تصمیم گیری درباره میزان، نوع و زمان مناسب سرمایه‌گذاری بستگی دارد. رشد روز افزون کمی و کیفی بازار سرمایه ایران طی سالیان اخیر، توجه بسیاری از فعالان اقتصادی، سرمایه‌گذاران و توده‌های مردم را به سرمایه‌گذاری در بازارهای مولد نظیر بورس اوراق بهادار جلب کرده است. با این وجود، سرمایه‌گذارانی که فاقد دانش و تجربه درباره عملیات بازارهای مالی هستند، ممکن است ثروت خود را در صورت سرمایه‌گذاری نادرست در اوراق بهادار از دست بدهند. مهمترین دغدغه برای سرمایه‌گذاران، تخصیص بهینه وجوه در دسترس به گزینه‌های مختلف سرمایه گذاری است که گزینه مذکور اهداف آنها را برآورده ساخته، ریسک آتی آنها را کمینه کرده و از یک درآمد معقول با رشد مناسب در آینده اطمینان حاصل نماید. بنابراین به موازات ورود سرمایه‌گذاران جدید که اغلب اطلاعات کافی درباره بازار سرمایه و تحلیل شرکت‌های بورسی ندارند، بهره‌گیری از ابزارهای مالی که به نحوی ریسک سرمایه گذاری را کاهش دهد، بیش از پیش احساس می‌شود. به خاطر ماهیت منحصر به فرد بازار سرمایه و اوراق بهادار و از آنجائی‌که سرمایه‌گذاری در این بازار کاری پیچیده و مستلزم دانش، مهارت و تخصص است، صندوق‌های سرمایه‌گذاری ابزار مناسبی برای سرمایه‌گذاران فراهم می‌کنند تا در پرتفویی متنوع از اوراق بهادار و دارای مدیریت حرفه‌ای با هزینه نسبتاً کم سرمایه‌گذاری کنند. یک صندوق سرمایه‌گذاری، ابزاری است که پس انداز عده‌ای از سرمایه‌گذاران را که هدف مالی مشترکی دارند، در قالب واحدها یک کاسه می‌نماید. وجوهی که از این طریق جمع‌آوری می‌شود، در ابزارهای مختلف بازار سرمایه نظیر سهام، اوراق قرضه، ابزارهای بازار پول و یا ترکیبی از این ابزارها سرمایه‌گذاری می‌شود (تاجبر، ۱۳۹۴). مزایای صندوق‌های سرمایه‌گذاری شامل مدیریت حرفه‌ای، مدیریت ریسک از طریق متنوع‌سازی، کاهش هزینه‌ها از طریق صرفه‌جویی به مقیاس، تقسیم‌پذیری و نقدشوندگی می‌باشد. با توجه به تعداد زیاد صندوق‌های سرمایه‌گذاری، عملکرد متفاوت آنها در گذشته و ویژگی-های متفاوت آنها، انتخاب صندوق سرمایه‌گذاری مناسب برای سرمایه‌گذاران امری حیاتی است. هدف از فرآیند انتخاب صندوق سرمایه‌گذاری انتخاب یک صندوق سرمایه‌گذاری از میان تعداد زیاد صندوق‌های سرمایه‌گذاری در دسترس با عنایت به ترجیحات سرمایه‌گذار، فضای اقتصادی و محدودیت‌هاست. در دیدگاه سنتی، ریسک و بازدهی گذشته صندوق‌های سرمایه‌گذاری تنها ویژگی‌هایی هستند که سرمایه‌گذاران را در انتخاب صندوق سرمایه‌گذاری کمک می‌کنند. بر اساس این دیدگاه، سرمایه‌گذاران عموماً صندوق سرمایه‌گذاری را بر مبنای

عملکرد گذشته آن انتخاب می‌کنند و توقع دارند که صندوق سرمایه‌گذاری عملکرد گذشته خود را تکرار کند. این در حالی است که عملکرد گذشته نمی‌تواند به تنهایی عملکرد آینده صندوق سرمایه‌گذاری را پیش‌بینی کند. لذا علاوه بر عملکرد گذشته شامل ریسک و بازده، عوامل متعدد ساختاری و رفتاری دیگری وجود دارد که می‌تواند بر بازدهی آتی صندوق سرمایه‌گذاری اثر گذارد. این عوامل می‌تواند شامل اندازه، هزینه‌های صندوق، حق الزحمه مدیریت، نسبت گردش پرتفولیوی، سبک مدیریت، جریان نقدینگی، ساختار مدیریت، عمر صندوق، میزان تمرکز، موقعیت مکانی، تجربه و مهارت مدیریت، جنسیت مدیریت، حجم معاملات، عمر مدیریت، ساختار حاکمیتی صندوق‌های سرمایه‌گذاری و ... باشد. علاوه بر این، در صورت اطلاع از عوامل اثرگذار بر تصمیم‌گیری در مورد انتخاب صندوق سرمایه‌گذاری، فرآیند تحلیل، تفسیر و انتخاب صندوق نیز موضوعی ناشناخته است که براساس خبرگی فرد تصمیم‌گیرنده صورت می‌پذیرد. بنابراین اگر بتوان این فرآیند را به شکل منطقی مدل کرد، امکان تصمیم‌گیری برای افراد غیره خبره نیز فراهم می‌گردد. با توجه به توضیحات فوق، مسأله اصلی این تحقیق علاوه بر تعیین ویژگی‌های مؤثر در عملکرد آینده صندوق‌های سرمایه‌گذاری در ایران ، طراحی یک مدل ترکیبی (هم از جنبه عوامل کمی و کیفی و روش تحقیق و هم از جنبه نظریه‌ها و علوم مورد استفاده) جهت تعیین وضعیت صندوق‌های سرمایه‌گذاری در آینده و ایجاد یک سیستم خبره تصمیم‌گیری در این حوزه می‌باشد.

از سوی دیگر، بیشتر مدل‌های تئوریک برای ارزیابی عملکرد و انتخاب صندوق‌های سرمایه‌گذاری نیز بر پایه مفروضات بازار کار است و بازارهای مالی نوظهور اغلب این مفروضات را برآورده نمی‌سازند. بازدهی در بازارهای نوظهور همراه با شرایطی نظیر نوسان‌پذیری بالا، هزینه معاملات بالا و حجم معاملات پایین است. به علاوه، داده‌هایی که از دنیای واقعی اخذ می‌گردند، ممکن است شامل تمامی انواع نویزها بوده و عدم قطعیت‌های بسیار و اطلاعات غیرکامل فراوانی داشته باشند که این امر در خصوص انتخاب صندوق‌های سرمایه‌گذاری نیز صادق است. لذا روش‌های سنتی در برخورد با این عدم قطعیت‌ها نیازمند اطلاعات اضافی است. بنابراین کار با حجم بالایی از داده‌ها توسط این سیستم‌ها مشکل است که تئوری‌هایی نظیر تئوری مجموعه راف[1] (RST) می‌تواند در جهت رفع این مشکل کمک‌کننده باشد. بنابراین، با توجه به این که بیشتر تحقیقات صورت گرفته در زمینه عملکرد، گروه‌بندی و انتخاب صندوق‌های سرمایه‌گذاری با توجه به عملکرد گذشته انجام گرفته و ویژگی‌های صندوق‌های سرمایه‌گذاری کمتر در نظر گرفته شده و مطالعه‌ای در رابطه با تمامی ویژگی‌های موثر در ایران انجام نگردیده، در این تحقیق کلیه ویژگی‌های قابل اندازه‌گیری صندوق‌های سرمایه‌گذاری موثر بر عملکرد آتی، شناسایی و مورد توجه قرار خواهد گرفت؛ همچنین با توجه به وجود محدودیت‌ها در اطلاعات و عدم قطعیت موجود در اطلاعات لازم است تا به کمک ترکیبی از علوم و نظریه‌های مرتبط، هم از بعد کیفی و هم از بعد ریاضی و کمی مورد بررسی قرار گیرد که در بعد کیفی می‌بایست ویژگی‌های موثر در عملکرد صندوق‌های سرمایه‌گذاری مشخص و در بعد کمی، تئوری‌های عدم قطعیت نظیر RST می‌تواند جهت طراحی موتور استنتاج در خصوص عملکرد صندوق‌های سرمایه‌گذاری براساس ویژگی‌ها بکار گرفته شوند.

۲- پیشینه پژوهش

هریل[۲] و همکاران (۲۰۱۰) در مطالعه‌ای با استفاده از مدل‌سازی مارکوف[۳] مبادرت به رتبه‌بندی صندوق‌های سرمایه‌گذاری نموده و نشان دادند پایداری ضعیفی در عملکرد آنها وجود دارد. نتیجه این تحقیق تأثیر عوامل و ویژگی‌های کیفی صندوق سرمایه‌گذاری بر عملکرد آن را با اهمیت توصیف نموده است. پیترز و چن و فرسون[۴] در ایالات متحده امریکا در سال ۲۰۱۰ به بررسی عملکرد ۱۴۰۰ شرکت و صندوق سرمایه‌گذاری آمریکایی بین سال‌های ۱۹۶۲ تا ۲۰۰۷ پرداختند. نتایج نشان داد که به طور کلی شرکت‌ها و صندوق‌های سرمایه‌گذاری آمریکایی در طول دوره مطالعه ثباتی در زمان‌بندی بازار از خود نشان نمی‌دهند. علاوه بر این عملکرد گذشته شرکت‌ها و صندوق‌های سرمایه‌گذاری آمریکایی کمکی به پیش‌بینی عملکرد آینده آنها نمی‌کند. همچنین شرکت‌ها و صندوق‌های سرمایه‌گذاری غیرفعال، توانایی کمتری در زمان‌بندی بازار نسبت به شرکت‌ها و صندوق-های سرمایه‌گذاری فعال دارند و از این رو بازده کمتری از شرکت‌های فعال دارند. بودینو و مارتنز[۵] (۲۰۱۰) مطالعه‌ای در رابطه با انتخاب صندوق سرمایه‌گذاری با استفاده از برخی ویژگی‌های صندوق‌های سرمایه‌گذاری انجام دادند. آنها به این نتیجه رسیدند که انتخاب صندوق سرمایه‌گذاری بر اساس عملکرد گذشته، توانمندی و نسبت گردش معاملات معنادار می‌باشد. اوانز و فالنبراک (۲۰۱۲)، صندوق‌های سرمایه‌گذاری که همه به صورت خرده‌فروشی و همه به صورت نهادی فروخته می‌شوند را با صندوق‌های سرمایه‌گذاری که فقط به صورت خرده فروشی فروخته می‌شوند مقایسه کردند. آنها دریافتند که جریان سرمایه سرمایه‌گذاران در مقایسه با سرمایه-گذاران خرد نسبت به نرخ‌های حق الزحمه و هزینه و عملکرد حساس می‌باشد. صندوق‌های سرمایه‌گذاری که به هر دو گروه واحد سرمایه‌گذاری می‌فروختند، نسبت به صندوق‌های کارگزاری که فقط خرده‌فروشی می‌کردند عملکرد بهتری داشتند. همچنین نتیجه گرفتند که نتیجه فوق به خاطر این نیست که سرمایه‌گذاران نهادی در جذب صندوق‌های سرمایه‌گذاری موفق‌تر می‌شوند، بلکه علت آن نظارت بهتر بر مدیران پرتفلیو توسط سرمایه-گذاران نهادی است. جباری و همکاران (۲۰۱۲) در مطالعه ای در رابطه با ارزیابی عملکرد صندوق‌های سرمایه-گذاری و رتبه‌بندی آنها با توجه به اندازه نمونه و اطلاعات ناکافی از تئوری سیستم خاکستری و درجه رابطه خاکستری[۶] استفاده نمودند. فریرا[۷] و همکاران (۲۰۱۳) با استفاده از مدل چهار عاملی کارهارت ویژگی‌هایی نظیر عمر، اندازه، حق الزحمه، ساختار مدیریت، مدت تصدی مدیریت و ویژگی‌های کشوری نظیر توسعه اقتصادی، توسعه مالی و حمایت از سرمایه‌گذار در نظر گرفتند. نتیجه این شد که صندوق‌های سرمایه‌گذاری بزرگ، صندوق‌های سرمایه‌گذاری با عمر کمتر، صندوق‌های سرمایه‌گذاری با حق‌الزحمه بالا و صندوق‌های سرمایه-گذاری با مدیران با تجربه‌تر عملکرد بهتری دارند. فانگ[۸] و همکاران (۲۰۱۴) در مطالعه‌ای یک چارچوب چند لایه برای ارزیابی چگونگی تاثیر سیستماتیک مشخصات مدیر بر عملکرد صندوق سرمایه‌گذاری ایجاد کردند، این ساختار شامل سه لایه از عناصر عملکردی است :۱) عملکرد جامع ۲) ریسک و بازده ۳) مهارت زمان‌سنجی و توانایی انتخاب سهام. نتیجه این بود که ویژگی‌های چندی وجود دارد که بر بازدهی ریسک و توانایی‌های مدیریت که بر عملکرد تاثیر دارد. بابالوس[۹] و همکاران (۲۰۱۵) با استفاده از روش تحلیل مرزی تصادفی[۱۰] عملکرد صندوق‌های سرمایه‌گذاری آمریکایی را بررسی کرد و به این نتیجه رسید سبک مدیریت و اندازه بر

عملکرد صندوق‌های سرمایه‌گذاری موثر است. روزگار[11] و همکاران (۲۰۱۵) با استفاده از مدل RST و تحلیل تشخیصی[12] صندوق‌های سرمایه‌گذاری ترکیبی[13] (ترکیب دارایی‌ها شامل سهام و اوراق با درآمد ثابت) را بر اساس ترکیب دارایی‌ها گروه‌بندی نمودند. نتایج حاصله نشان داد عملکرد مدل RST نسبت به تحلیل تشخیصی بهتر بوده و با آزمون مدل با اطلاعات واقعی نتایج امیدوارکننده نمایان شد. قمر و سینف (۲۰۱۶) از روش تحلیل پوششی داده‌ها برای پیش‌بینی عملکرد ۴۶ صندوق در سال‌های ۲۰۱۵-۲۰۰۶ استفاده کردند که نتایج آنها نشان داد که در انتهای دوره صندوق‌ها عملکرد خوبی ندارند اما در دریافت عمل باثبات و پایدار هستند. آنها دریافتند که این روشی در رتبه‌بندی صندوق‌ها و پیش‌بینی عملکرد آنها در تصمیم‌گیری سرمایه‌گذاران مفید است. خلیل و همکاران (۲۰۱۶) در مطالعه‌ای در خصوص اثر ویژگی‌های مدیریتی نظیر سن مدیر، کیفیت مدیر، تجربه مدیر عمر صندوق و حق‌الزحمه مدیریت بر عملکرد صندوق‌های سرمایه‌گذاری در پاکستان با استفاده از روش تجزیه و تحلیل رگرسیونی و با استفاده از روش پنل دیتا (اثرات ثابت) انجام دادند. نتایج نشان داد که کیفیت مدیریت، سن و تجربه در صنعت صندوق‌های سرمایه‌گذاری به طور مثبت و معنادار با عملکرد صندوق رابطه دارد در حالی که عمر صندوق به طور منفی با عملکرد در رابطه است حق الزحمه مدیریت با عملکرد رابطه ندارد.

شینی زاده عمادی (۱۳۹۳) به مطالعه رابطه بین نحوه مدیریت صندوق های سرمایه گذاری مشترک و عملکرد آنها با استفاده از مدل چهار عاملی کاهارت پرداخت تا تأثیر مدیریت صندوق‌ها را بر عملکرد صندوق مورد سنجش قرار دهد. نتایج نشان داد که مازاد بازده صندوق با عامل صرف ریسک بازار، عامل ارزش دفتری به ارزش بازار و عامل روند رابطه مثبت و معنادار دارد، همچنین عامل اندازه و انحراف معیار اختلاف بازده صندوق و بازار، رابطه منفی با مازاد بازده صندوق داشت. صادقی مقدم و همکاران (۱۳۹۵)، با بهره‌گیری از روش‌های تصمیم‌گیری چندمعیارهٔ TOPSIS، VIKOR و تکنیک مشابهت تلاش کردند، توانایی روش‌های یادشده در رتبه‌بندی صندوق‌های منتخب از میان صندوق‌های مشترک در بازار سرمایهٔ ایران را بیازمایند و از سوی دیگر در ارزیابی صندوق‌ها، سه دسته شاخص‌های عمومی ارزیابی عملکرد صندوق‌ها (عمر، ارزش خالص هر واحد سرمایه‌گذاری، خالص ارزش دارایی‌ها، درصد دارایی‌های نقدی و درصد بازدهی)، شاخص‌های ارزیابی عملکرد تعدیل‌شده با ریسک (شارپ، ترینر و جنسن) و شاخص‌های ارزیابی عملکرد تعدیل‌شده با ریسک با لحاظ نیم‌واریانس (شارپ و ترینر تعدیل‌شده و جنسن تعدیل‌شده با نیم بتا) را به صورت مجزا و نیز با هم را بررسی کردند تا توان این سه رویکرد در ارزیابی صندوق‌ها با شاخص‌های مختلف مقایسه شود. نتایج نشان داد که نتایج حاصل از روش‌های TOPSIS، VIKOR و مشابهت در ارزیابی صندوق‌های سرمایه‌گذاری مشترک رفتارهای متفاوتی از خود نشان داده است. از دیدگاه شاخص‌ها برای نتایج به دست آمده، درصد داراییهای نقدی، خالص ارزش دارایی‌های صندوق، NAV و عمر صندوق در میان رتبه‌های برتر متفاوت است.

۳- تئوری مجموعه‌های راف[۱۴] (RST)

تئوری مجموعه راف قادر به تحلیل جدول داده‌هاست. داده‌هایی که می‌تواند از طریق اندازه‌گیری و یا خبرگان حاصل شده باشد. رویکرد RS، الگوریتمی کارا برای یافتن الگوهای پنهان در داده‌ها، کمینه‌نمودن مجموعه داده‌ها (تقلیل داده‌ها)، ارزیابی معناداری داده‌ها و ایجاد مجموعه‌ای از قوانین تصمیم‌گیری برای داده‌ها ارائه می‌نماید. در واقع در این تئوری، مفاهیم مبهم با دو مفهوم زوجی دقیق برای آن مفهوم مبهم، به نام تقریب پایین و بالا جایگزین می‌شوند. تقریب پایین شامل تمام اجزاء که بطور حتم به آن مفهوم تعلق دارند، شده و تقریب بالا شامل تمام اجزاء که امکان تعلق به آن مفهوم را دارند، می‌شود. تفاوت حد بالا و پایین، ناحیه مرزی مفهوم مبهم را می‌سازد. نقطه آغاز RST مجموعه داده‌هاست که معمولاً در قالب یک جدول سازمان داده شده و تحت نام سیستم اطلاعاتی یا پایگاه داده معرفی می‌شوند. اصلی‌ترین عملیات در RST تقریب بالا و پایین است و این تقریب‌ها برای تشخیص اجزای کاملا وابسته و یا نیمه وابسته در جدول اطلاعات بکار می‌روند. تقلیل داده[۱۵] نیز یکی از مهم‌ترین مفاهیم این تئوری است. به شکل ریاضی زوج $S=(U,A)$ را یک سیستم اطلاعاتی می‌نامند که U یک مجموعه غیرتهی و محدود از اعضا بوده و به آن مجموعه مرجع گویند و A یک مجموعه غیرتهی و محدود از شاخص‌هاست، بطوری که برای هر $a\in A$ داریم $a:U\to V_a$ که V_a را مجموعه ارزش[۱۶] a می‌نامند. به بیان دیگر $S=(U,A,V,f)$ یک سیستم اطلاعاتی است که $U=\{U_1,U_2,...,U_{|U|}\}$ مجموعه مرجع و $A=\{a_1,a_2,...,a_{|A|}\}$ مجموعه شاخص‌هاست. $V=\cup V_a$ و V_a دامنه شاخص a است و $f:U\times A\to V$ تابعی است که برای $\forall a\in A,\forall x\in u$ ارزش شاخص هر یک از اجزا را بصورت $f(x,a)\in V_a$ تخصیص می‌دهد (کریمی و صادقی‌مقدم، ۱۳۹۵). اگر یک سیستم اطلاعاتی به شکل $S=(U,A,V,f)$ داشته باشیم آنگاه شاخص‌های موجود در مجموعه A را می‌توان به دو زیر مجموعه جدا از هم تقسیم نمود که به یک گروه مشخصه‌های موقعیتی یا C گوییم و به گروه دیگر شاخص‌های تصمیم یا D و خواهیم داشت $A=C\cup D$, $C\cap D=\phi$ در اینصورت S را جدول تصمیم می‌نامیم. یک سیستم اطلاعاتی ممکن است بی دلیل بزرگ شده و به دو علت دچار تکرار شود. اول به دلیل اعضای مشابه و غیر قابل تشخیص از یکدیگر و دوم به علت وجود برخی شاخص‌های زائد در سیستم اطلاعاتی. سیستم اطلاعاتی $S=(U,A)$ را در نظر بگیرید که برای هر $B\subseteq A$ یک رابطه هم ارزی به شکل زیر برقرار است:

$$IND_S(B)=\{(x,x')\in U^2:\forall a\in B,\ a(x)=a(x')\}$$

اگر $(x,x')\in IND_S(B)$ باشد آنگاه با در نظر گرفتن شاخص های B اعضای x , x' غیر قابل تشخیص از یکدیگر می‌باشند. به بیان دیگر این دو عضو را نمی توان از طریق شاخص‌های موجود در مجموعه B از هم متمایز نمود. کلاس‌های هم ارز مربوط به رابطه IND را به شکل $[x]_B$ نشان می‌دهند. اندیس S را زمانی که سیستم اطلاعاتی مشخص است می‌توان حذف نمود. پس به طور خلاصه می‌توان با توجه به هر زیر مجموعه از

شاخص‌ها (B) اعضای غیر قابل تمیز[۱۷] از هم را پیدا نمود. فرض اساسی در RST این است که دانش را می‌توان طبقه‌بندی نمود. به بیان دیگر،این تئوری به عنوان یک چارچوب استاندارد برای یافتن واقعیت از درون داده-هاست. RST دانش بدست آمده ودسته بندی شده را به عنوان بخشی از مجموعه داده ها در نظر می گیرد. براساس دانش موجود، در خصوص تعلق $x \in U$ به مجموعه $X \subseteq U$ سه حالت به شکل زیر وجود دارد:

۱) عضو x مطلقا به مجموعه X تعلق دارد

۲) عضو x مطلقا به مجموعه X تعلق ندارد

۳) عضو x ممکن است به مجموعه X تعلق داشته باشد و ممکن است تعلق نداشته باشد.

براساس تحلیل بالا دو مفهوم بسیار مهم در تئوری RS به نام تقریب پایین و تقریب بالا[۱۸] مطرح می‌شود.

تعریف: اگر سیستم اطلاعاتی $S = (U, A)$ را داشته باشیم و $B \subseteq A$, $X \subseteq U$ می‌توانیم X را تنها با اطلاعات موجود در B و با ساختن تقریب پایین و بالای B از X تقریب بزنیم که داریم:

$$\underline{BX} = \underline{apr}_B(X) = \{x : [x]_B \subseteq X\}$$

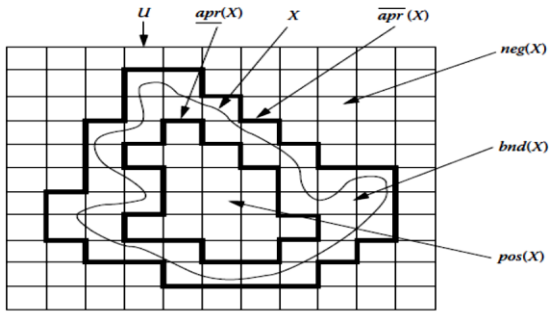
$$\overline{BX} = \overline{apr}_B(X) = \{x : [x]_B \cap X \neq \phi\}$$

اعضای $\underline{BX}$ را می‌توان با قطعیت و برپایه دانش موجود در B عضو X دانست در حالی که اعضا $\overline{BX}$ را می-توان اعضای محتمل یا ممکن X دانست.

تعریف: مجموعه $bnd_B(X) = \overline{BX} - \underline{BX}$ را ناحیه مرزی B از X می‌نامند و شامل آن دسته از اعضا است که نمی‌توانیم براساس دانش حاصل از B با قطعیت جز X دسته‌بندی نماییم.

تعریف: به مجموعه $U - \overline{BX}$ را ناحیه خارج B از X می‌نامیم و شامل اعضایی است که بطور قطع می‌دانیم عضو X نیستند، به عبارت دیگر $neg_B(X) = \bigcup \{x : [x]_B \cap X = \phi\}$.

به مجموعه‌ای rough گفته می‌شود که ناحیه مرزی آن تهی نباشد. تقریب پایین، تقریب بالا، ناحیه متمم[۱۹] و حدود مرزی[۲۰] مجموعه X در شکل زیر نشان داده شده است:



شکل ۱. نمایش مفاهیم مربوط به RST

تعیین ویژگی‌های مهم سیستم اطلاعاتی براساس رابطه «غیر قابل تمیز» امکان‌پذیر است. اگر تعداد اعضای همسانی که از مجوعه شاخص‌های A استخراج می‌شود با تعداد اعضای همسان حاصل از $A-a_i$ برابر باشد، آنگاه شاخص a_i را به عنوان شاخص زائد یا تکراری[۲۱] می‌نامند و در غیر این صورت به آن شاخص ضروری[۲۲] گفته می‌شود. یکی از موضوعات مهم در تحلیل داده‌ها، کشف وابستگی بین شاخص‌هاست. زمانی یک مجموعه از شاخص‌ها مانند D کاملا وابسته به مجموعه‌ای دیگر از شاخص‌ها مانند C است که تمامی مقادیر شاخص‌های D بصورت یکتا توسط مقادیر شاخص‌های C تعیین گردد؛ در این صورت داریم $C \Rightarrow D$. به بیان دیگر D را کاملاً وابسته به C می‌نامیم اگر یک تابع وابستگی بین مقادیر D و C وجود داشته باشد. اگر D و C دو زیر مجموعه از A باشند آنگاه D را با شدت $k \ (0 \le k \le 1)$ وابسته به C می‌دانیم و به شکل $C \Rightarrow_k D$ نشان می‌دهیم که مقدار k برابراست با:

$$k = \gamma(C,D) = \frac{|POS_C(D)|}{|U|} \quad , \quad POS_C(D) = \bigcup_{X \in U/D} \underline{C}(X)$$

اگر $k=1$ باشد D کاملا به C وابسته است و در غیر اینصورت وابستگی نسبی و با شدت k بین این دو مجموعه وجود دارد. در یک سیستم اطلاعاتی به شکل $S=(U,A,V,f)$, $p \subseteq A$, $X \subseteq U$ می‌توان دقت تخمین X را به شکل زیر تعریف نمود:

$$\alpha_p(X) = \frac{\left|\underline{apr}_p(X)\right|}{\left|\overline{apr}_p(X)\right|}$$

دقت تخمین نشان‌دهنده میزان فهم حاصل از X و براساس دانش موجود است. روشن است که $0 \le \alpha_p(X) \le 1$. اگر $\alpha_p(X)=1$ باشد آنگاه مجموعه X یک مجموعه کریسپ بوده و اگر $\alpha_p(X) \prec 1$ باشد، آنگاه مجموعه X

یک مجموعه راف می‌باشد. فرض کنید که $X=\{X_1,X_2,...,X_n\}$ یک افراز از مجموعه مرجع U است که $X_i(i=1,2,...,n)$ یک کلاس از X است و $p \subseteq A$ آنگاه کیفیت تخمین X به شکل معادله زیر تعریف می‌شود:

$$\gamma_p(X)=\frac{\sum_{i=1}^{n}\left|\underline{apr}_P(X_i)\right|}{|U|}$$

اگر کیفیت تقریب $\gamma_p(X)=1$ باشد آنگاه X کاملا وابسته به P است. اگر $0 \prec \gamma_p(X) \prec 1$ باشد آنگاه می‌توان گفت که X تا حدودی به P وابسته است و این نشان می‌دهد که تنها بخشی از شاخص‌ها در P قابل دسترسی هستند، یا به بیان دیگر مجموعه داده‌ها دارای نقصان است. هر چقدر وابستگی بیشتر باشد اهمیت P بیشتر است . اگر تعداد اعضای مجموعه پس از حذف هر شاخص کم شود نشان‌دهنده وابستگی مجموعه به آن شاخص است (کریمی و صادقی‌مقدم، ۱۳۹۵). کاهش شاخص‌ها موجب می‌شود که تنها آن دسته از شاخص‌ها که تفکیک و تقریب را باعث می‌گردند، حفظ شوند. شاخص‌های کنار گذاشته‌شده زائد بوده، چرا که حذف آنها نمی‌تواند دسته‌بندی را خدشه‌دار نماید. معمولاً چندین زیر مجموعه با این ویژگی‌ها وجود دارد که به آندسته از زیرمجموعه‌ها که دارای حداقل شاخص هستند بی‌زائده گویند. هسته[۲۳] و بی‌زائده[۲۴] دو مفهوم اساسی هستند که بی‌زائده کوچکترین زیر مجموعه از شاخص‌هاست که غیروابسته بوده و قدرت تقسیم‌بندی آن معادل کل شاخص‌هاست. لذا بخش حیاتی و ضروری سیستم اطلاعات را تشکیل می‌دهد و هسته، قسمت موجود و مشترک در بین تمامی بی‌زائده‌ها را می‌نامند. در صورتی که $S=(U,A,V,f)$ و I یک رابطه غیرقابل تمیز در U باشد، اگر $p \subseteq A$, $Q \subseteq A$ آنگاه وابستگی بین شاخص‌ها را می‌توان به شکل زیر بیان نمود:

۱) اگر $I(p) \subseteq I(Q)$ باشد آنگاه مجموعه شاخص‌های Q به مجموعه شاخص‌های P وابسته است، یعنی $p \Rightarrow Q$.

۲) اگر $p \Rightarrow Q$ و $Q \Rightarrow p$ در اینصورت مجموعه شاخص‌های P معادل مجموعه شاخص های Q است یعنی $p \Leftrightarrow Q$.

وابستگی بین شاخص‌ها یک مفهوم بسیار مهم در کاربردهای عملی از جدول تصمیم است. اگر مقدار تمامی شاخص‌های مجموعه D بطور یکتا با مقادیر شاخص‌های مجموعه C قابل تعیین باشد، می‌گوییم D کاملا وابسته به C است واگر تنها برخی مقادیر شاخص D از طریق مقادیر شاخص C قابل درک باشد، می‌گوییم یک وابستگی عملکردی نسبی بین C و D وجود دارد. در واقع وابستگی عملکردی نشان می‌دهد تا چه حد دانش C می‌تواند به ساختن دانش D منجر شود. در RST کشف وابستگی در تحلیل دانش بسیار مهم است. کیفیت کلاسه‌بندی مجموعه شاخص‌های بی‌زائده با کیفیت مجموعه شاخص‌های اصلی برابر است. اگر حداقل شاخص‌های مجموعه $P \subseteq C \subseteq A$ دارای $\gamma_P(X)=\gamma_C(X)$ باشد، آنگاه مجموعه P را بی‌زائده» مجموعه C می‌نامند و به شکل

RED(P) نشان می‌دهند. به بیان ساده‌تر بی‌زائده کوچکترین زیر مجموعه از شاخص‌های موقعیتی است که شاخص‌های تکراری یا زائد ندارد، اما منجر به کلاسه‌بندی درست می‌شود. کشف وابستگی شاخص‌ها منجر به تولید بی‌زائده‌هایی می‌شود که دارای کیفیت تقسیم‌بندی نظیر مجموعه اصلی هستند. از آنجا که ممکن است بیش از یک بی‌زائده در جدول تصمیم وجود داشته باشد، لذا نقطه مشترک بین تمامی بی‌زائده‌ها را هسته جدول تصمیم می‌نامند که به شکل زیر نشان داده می‌شود:

$$CORE(P) = \bigcap_{R_i \in RED(P)} R_i \qquad i = (1,2,...,n)$$

هسته، مهم‌ترین مجموعه شاخص‌های یک سیستم اطلاعاتی را مشخص کرده و ممکن است یک مجموعه تهی باشد (کریمی و صادقی‌مقدم، ۱۳۹۵). شاخص‌ها در یک سیستم اطلاعاتی دارای اهمیت یکسان نمی‌باشند و برخی از این شاخص‌ها را می‌توان حذف نمود بدون آنکه اطلاعاتی از دست برود. توضیح مفهوم و ایده کاهش شاخص‌ها را می‌توان با تشریح مفهوم معناداری شاخص‌ها[۲۵] شروع نمود، چرا که این مفهوم امکان ارزشیابی شاخص‌ها را فراتر از تقسیم آنها به دو دسته ضروری و غیرضروری[۲۶] فراهم کرده و از طریق تخصیص یک عدد حقیقی در بازه $[0,1]$ به هر شاخص درجه اهمیت شاخص را در سیستم اطلاعاتی مشخص می‌کند. معناداری شاخص a در جدول اطلاعاتی $S = (U, CUD)$ را می‌توان با سنجش تأثیر حذف یک شاخص $a \in C$ از مجموعه شاخص C ارزشیابی نمود. همانگونه که قبلا گفته شد عدد $\gamma(C,D)$ بیان‌کننده درجه وابستگی بین شاخص‌های C و D است. به بیان دیگر دقت تقریب U/D به کمک C را نشان می‌دهد. سوال اینست که مقدار ضریب $\gamma(C,D)$ با حذف یک شاخص مانند a چگونه تغییر می‌کند. به بیان دیگر تفاوت بین $\gamma(C,D)$ و $\gamma(C-\{a\},D)$ چیست. ما این تفاوت را نرمالایز کرده و معناداری یک شاخص مانند a را به شکل زیر تعریف می‌کنیم:

$$\sigma_{(C,D)}(a) = \frac{\left(\gamma(C,D) - \gamma(C-\{a\},D)\right)}{\gamma(C,D)} = 1 - \frac{\gamma(C-\{a\},D)}{\gamma(C,D)}$$

بنابراین ضریب $\sigma_{(C,D)}(a)$ را می‌توان مقدار خطای دسته‌بندی در صورت حذف شاخص a نامید. ضریب معناداری را می‌توان به یک مجموعه از شاخص‌ها به شکل معادله زیر تعمیم داد:

$$\sigma_{(C,D)}(B) = \frac{\left(\gamma(C,D) - \gamma(C-B,D)\right)}{\gamma(C,D)} = 1 - \frac{\gamma(C-B,D)}{\gamma(C,D)}$$

که B یک زیر مجموعه از C می‌باشد. اگر B یک بی‌زائده برای مجموعه C باشد، آنگاه $\sigma(C-B) = 0$ خواهد شد. هر زیر مجموعه B از مجموعه C می‌تواند به عنوان یک بی‌زائده تقریبی برای C باشد و مقدار عدد زیر:

$$\varepsilon_{(C,D)}(B) = \frac{\left(\gamma(C,D) - \gamma(B,D)\right)}{\gamma(C,D)} = 1 - \frac{\gamma(B,D)}{\gamma(C,D)}$$

که بطور خلاصه با $\varepsilon(B)$ نشان داده می‌شود را مقدار خطای تقریب بی‌زائده[۲۷] می‌نامند و بیان‌کننده آنست که با چه دقتی مجموعه شاخص B می‌تواند مجموعه شاخص‌های C و D را تقریب بزند. مفهوم تقریب بی زائده عمومیت یافته مفهوم بی زائده است، چرا که زیرمجموعه B با حداقل تعداد شاخص را که $\gamma(C,D)=\gamma(B,D)$ باشد و یا $\varepsilon_{(C,D)}(B)=0$ را بی‌زائده می‌نامند. در بسیاری از سیستم‌های یادگیری استنتاج قوانین یکی از وظایف ضروری است. مجموعه تمامی شاخص‌های موقعیتی را کلاس‌های موقعیتی[۲۸] S نامیده و به شکل $X_i(i=1,2,...,n)$ نشان می‌دهند. مجموعه شاخص‌های تصمیم را کلاس‌های تصمیم[۲۹] S نامیده و به شکل $Y_j(j=1,2,...,n)$ نشان می‌دهند. قوانین تصمیم (C,D) را به شکل $\{r_{ij}\}$ نشان می‌دهیم. برای $\vee i,j\ ,\ if\ \ X_i \subseteq Y_j$

آنگاه قانون r_{ij} قطعی است و در غیر اینصورت غیرقطعی است. نمایش یک قانون بصورت زیر است:

$$IF\ ,\ f(x,q_1)=r_{q1}\ \wedge\ f(x,q_2)=r_{q2}\ \wedge ...\wedge f(x,q_p)=r_{qp}\ \ THEN\ \ x\in Y_{j1}\vee Y_{j2}\vee ...\vee Y_{jk}$$
$$\{q_1,q_2,...,q_p\}\subseteq C\ ;\ \ (r_{q1},r_{q2},...,r_{qp})\in V_{q1}\times V_{q2}\times ...V_{qp}\ .$$

قسمت IF یک قانون را موقعیت و قسمت THEN را تصمیم می‌نامند. اگر یک عضوهم از نظر موقعیت و هم از نظر تصمیم سازگار باشد، نتایج یکنوا خواهند بود. تعداد عضوهایی که با یک قانون مطابقت دارند را پشتیبان[۳۰] آن قانون می‌نامند. تولید قوانین تصمیم از یک جدول کار پیچیده ای است و راه حل‌های پیشنهادی برای آن می‌تواند یکی از تاکتیک‌های زیر باشد:

۱) تولید مجموعه‌ای کمینه از قوانین که تمامی عضوهای یک جدول تصمیم را پوشش دهد.
۲) تولید یک مجموعه جامع از قوانین که شامل تمامی قوانین ممکن از یک جدول باشد.
۳) تولید یک مجموعه قوی از قوانین یا یک مجموعه نسبتاً قابل تشخیص که هر قانون اعضای نسبتاً زیادی را پوشش دهد و الزاما تمامی اعضا را شامل نشود.

فرآیند مدل‌سازی راف

فرآیند مدل‌سازی راف برای این تحقیق به به سه فاز تقسیم می‌گردد:

فاز پیش پردازش: در این فاز داده‌های ناقص تکمیل، داده‌های موجود اصلاح، شاخص‌ها انتخاب و داده‌ها نیز گسسسته می‌شوند. همچنین تقسیم‌بندی داده‌ها به دو گروه آموزشی و آزمایشی که بصورت تصادفی انجام می‌پذیرد، در این مرحله صورت می‌گیرد.

فاز تحلیل و تولید قوانین: این فاز شامل تولید دانش اولیه از طریق محاسبه «بی‌زائده»، استخراج قوانین از درون بی‌زائده‌ها، ارزیابی قوانین و فرآیند پیش‌بینی می‌شود.

فاز کلاسه‌بندی و پیش‌بینی: در این فاز از قوانین استخراج شده در مرحله قبل برای پیش‌بینی مشاهدات آتی استفاده می شود (کریمی و صادقی‌مقدم، ۱۳۹۵).

در ادامه هر یک از گام‌ها به تفصیل توضیح داده شده و نتیجه محاسبات حاصل از آن ارائه گردیده است.

- **تکمیل داده‌های بدون مقدار و محاسبات لازم**

در دنیای واقعی معمولا مقدار برخی داده‌ها مشخص نیست. این داده‌ها ممکن است بر روی نتایج حاصل از تکنیک‌های تئوری مجموعه‌های راف اثرگذار باشد. لذا هدف از فرآیند تکمیل داده‌ها حذف یا استفاده از سایر تکنیک‌های موجود جهت برخورد با مقادیر نامعلوم است. در نرم افزار *ROSETTA* روش‌های مختلفی برای تکمیل داده‌های بدون مقدار در نظر گرفته شده است که شامل:

- حذف داده‌های ناقص[۳۱]: به کمک این الگوریتم ، مشاهداتی که دارای مقادیر جا افتاده (خالی) در جدول اطلاعاتی هستند حذف گردیده و در محاسبات اثرگذار نخواهند بود.
- میانگین و میانه[۳۲]: با انتخاب این گزینه داده‌های جا افتاده مر بوط به شاخص‌های عددی با میانگین مشاهدات برای آن شاخص جایگزین می‌شود. همچنین برای شاخص‌های رشته‌ای از مد مشاهدات مربوط به آن شاخص برای پر کردن داده‌های جا افتاده استفاده می‌شود.
- میانگین و میانه شرطی[۳۳]: مانند مورد بالا بوده، با این تفاوت که برای محاسبه میانگین و مد به شاخص تصمیم‌گیری نیز توجه کرده و در محاسبه تنها داده‌های مربوط به کلاس‌های تصمیم مشابه را می-گنجاند.
- تکمیل ترکیبی[۳۴]: در این گزینه داده‌های جا افتاده با مجموعه‌ای از داده‌های موجود و ممکن پر می‌شود. به بیان دیگر یک گزینه به چندین گزینه تبدیل می‌شود که داده‌های جا افتاده آن با داده‌های ممکن پر شده است. در استفاده از این روش باید توجه نمود که تعداد ترکیب‌های ممکن با افزایش تعداد داده‌های جا افتاده به شدت افزایش می‌یابد.
- تکمیل ترکیبی شرطی[۳۵]: در این حالت نیز مانند حالت قبل از ترکیب داده‌های ممکن استفاده می‌شود، با این تفاوت که کلاس تصمیم نیز در تعیین داده‌های ممکن اثر گذار است. به بیان دیگر با در نظر گرفتن ویژگی‌هایی که دارای مقدار مشابه در گزینه تصمیم هستند داده‌های ممکن استخراج می‌شوند.

در این تحقیق برای تحلیل‌ها هریک از روش‌های فوق به تفکیک اجرا گردید که با توجه به حجم بالای محاسبات مربوط به روش‌های تکمیل ترکیبی و عدم تمایز محسوس بین نتایج، تنها روش حذف داده‌های ناقص، روش میانگین و میانه شرطی در تحلیل‌ها مورد استفاده قرار گرفته است.

- **گسسته نمودن داده‌ها**

زمانی که هدف از تحلیل داده ها، کلاسه‌بندی و پیش‌بینی است واضح است که درجه اهمیت هر یک از شاخص-ها باید مورد توجه قرار گیرد. در چنین حالت‌هایی معمولاً از اطلاعات کمکی در خصوص وزن هر یک از شاخص-

ها استفاده می‌شود. اما در رویکرد مجموعه‌های راف هیچگونه اطلاعات اضافی‌تر از آنچه که در جدول داده‌ها موجود است مورد نیاز نمی‌باشد. اصولا تئوری راف سعی دارد تا از داده‌های موجود در سیستم اطلاعاتی وزن و قدرت هریک از شاخص‌ها را در کلاسه‌بندی مشخص نماید. بنابراین از برخی استراتژی‌ها برای گسسته نمودن ارزش‌های واقعی استفاده می‌کند که کیفیت الگوریتم‌های یادگیری به این استراتژی‌ها وابسته است. این الگوریتم‌ها به دنبال یافتن برش‌هایی برای تعیین بازه‌ها هستند. تمامی مقادیری که بین هر بازه هستند به یک مقدار مشخص تبدیل می‌شوند و در نتیجه شاخص‌های عددی به شاخص‌هایی که می‌توانند دسته‌بندی شوند تبدیل می‌گردند. در این استراتژی‌ها، فرآیندهای تبدیل داده‌ها و یافتن برش‌های مناسب بر روی داده‌ها جهت تعیین بازه‌ها مورد استفاده قرار گرفته و به تمامی مقادیر یک بازه ارزش برابر داده می‌شود. این کار باعث کاهش تعداد مقادیر شاخص‌ها شده و منجر به تولید قوانین مشخص‌تر می‌شود. تنها شاخص‌های فعال در این الگوریتم‌ها وارد می‌شوند، لذا شاخص‌های موقعیتی رشته‌ای را می‌توان موقتاً غیرفعال نمود و پس از استفاده از الگوریتم‌های گسسته‌کردن، مجدد آنها را فعال نمود. الگوریتم‌های گسسته‌کردن را می‌توان در قالب سه دسته به شرح زیر تقسیم نمود:

۱) هر شاخص بصورت مستقل در نظر گرفته شده و هیچ اطلاعاتی از شاخص‌های تصمیم در فرآیند دخالت نکند. این الگوریتم‌ها را تک متغیری و بدون نظارت[۳۶] می‌نامند.
۲) تنها یک شاخص موقعیتی در یک زمان ولی با در نظر گرفتن شاخص تصمیم مورد توجه قرار می گیرد که به آنها الگوریتم‌های تک متغیری و با نظارت[۳۷] می‌گویند.
۳) تمامی شاخص‌های موقعیتی بطور همزمان در نظر گرفته شده و به شاخص‌های تصمیم نیز توجه می‌شود که الگوریتم‌های چند متغیری و با نظارت[۳۸] نامیده می‌شوند.

الگوریتم‌های مورد استفاده برای گسسته‌کردن داده‌ها در ادامه توضیح داده شده‌اند:

- الگوریتم بولین[۳۹]: این الگوریتم از تابع بولین استفاده کرده ونتیجه حاصل از آن کوچکترین مجموعه از شاخص‌هاست که قدرت تمیزدهی برابر مجموعه اصلی دارد.
- گسسته‌کردن دستی[۴۰]: این گزینه به کاربران اجازه می‌دهد تا بصورت دستی برش‌های لازم برای دسته‌بندی ویژگی‌ها را مشخص نمایند.
- الگوریتم آنتروپی[۴۱]: این الگوریتم براساس بخش‌بندی برگشتی[۴۲] مجموعه ارزش‌های هر یک از شاخص‌ها پایه‌ریزی شده است، بطوری که شاخص آنتروپی بهینه گردد. در این الگوریتم اگر هیچ برشی برای شاخص پیدا نشود آن شاخص بدون تغییر باقی می‌ماند.
- بازه‌های یکسان[۴۳]: این الگوریتم از دسته تک متغیری بدون نظارت بوده و یک مقدار ثابت n را در نظر گرفته و هیستوگرام هر شاخص را نمایش هد. n-1 برش بطوری تنظیم می‌کند که تقریبا تعداد برابری از گزینه‌ها در هریک از n بازه قرار گیرند. در نتیجه n-1 برش ایجاد می‌شود بطوری که فضای بین دو برش همسایه در هیستوگرام نرمال نزدیک به $1/n$ باشد.

- الگوریتم ساده[۴۴]: در این گزینه از یک الگوریتم هیورستیک ساده استفاده می‌شود که ممکن است منجر به تعداد زیادی برش شود. در بدترین حالت به هر ارزش مشاهده شده بازه منحصر به فردی تعلق می‌گیرد.
- به منظور سادگی فرض کنید تمامی شاخص‌های موقعیتی A عددی هستند. برای هر شاخص موقعیتی a می‌توانیم ارزش‌های V_a آنرا مرتب کنیم. C_a مقدار میانه ارزش دو گزینه برای آن شاخص است.
- الگوریتم شبه ساده[۴۵]: این الگوریتم نیز مانند بالاست با این تفاوت که مقدار شاخص تصمیم را نیز در نظر می‌گیرد.

در این پژوهش از الگوریتم‌های بولین، آنتروپی، یکسان و نیو برای گسسته کردن داده‌ها استفاده گردیده است.

- **تولید بی‌زائده‌ها**

یکی از مهمترین جنبه‌ها در تحلیل جدول اطلاعاتی، استخراج و حذف شاخص‌های تکراری و همچنین شناسایی مهمترین شاخص از درون داده‌هاست. شاخص‌های تکراری به شاخص‌هایی گفته می‌شود که حذف آنها هیچ تاثیری بر روی شدت وابستگی بین شاخص‌های موجود و تصمیم ندارد. شدت وابستگی اندازه‌ایست که قدرت تمیز بین مشاهدات را نشان می‌دهد. حداقل زیرمجموعه شاخص‌ها که شدت وابستگی آنها حداکثر است را بی‌زائده می‌نامند. محاسبه هسته و بی‌زائده‌های یک جدول اطلاعاتی در واقع انتخاب شاخص‌های مهم و مرتبط آن سیستم اطلاعاتی است.

همانطور که بیان گردید برای بدست آوردن «بی‌زائده» می‌توان از «تابع تمیز یا تشخیص» استفاده نمود که این تابع از ماتریس تمیز حاصل شده است. لازم به ذکر است که می‌توان یک تابع تمیز را از همه ورودی‌های موجود در ماتریس تمیز ساخت که در این صورت بی‌زائده حاصله قادر به تمیز همه مشاهدات از همدیگر و با گروه‌های تصمیم مختلف می‌باشد. به این گونه بی‌زائده‌ها *full-reducts* می‌گویند. همچنین می‌توان تنها بی-زائده‌هایی را بدست آورد که امکان تفکیک مشاهدات موجود را داشته باشند که به آنها *object-related-reducts* گفته شده و مرسوم‌ترند. همچنین می‌توان در ماتریس، مشاهدات موجود در گروه‌های تصمیم مشابه را نیز وارد نمود تا با این کار بی‌زائده‌هایی حاصل شود که قادر به تمیز مشاهدات با گروه‌های تصمیم مشابه باشد، در غیر این صورت ماتریس ساخته شده را *modulo Decision* می‌نامند که مرسوم‌ترند.

هر بی‌زائده در مجموعه بی‌زائده‌های تولید شده دارای یک شاخص پشتیبان[۴۶] است. این شاخص نشان قدرت آن بی‌زائده بوده و ممکن است براساس الگوریتم مورد استفاده متفاوت تفسیر گردد. مقدار این شاخص برابر نسبت درست نمایی آن بی‌زائده ضربدر ۱۰۰ می‌باشد.

متدهای موجود برای محاسبه بی‌زائده‌ها به شرح زیر می‌باشد:

الگوریتم ژنتیک: در این روش از الگوریتم ژنتیک برای محاسبه حداقل مجموعه شاخص‌ها استفاده می‌شود و می-توان پارامترهای مربوط به الگوریتم ژنتیک را تنظیم نمود.

الگوریتم جانسون[۴۷]: این گزینه شامل الگوریتم‌های ساده برای محاسبه بی‌زائده‌های منفرد است که توسط جانسون ارائه شده است.

در این بررسی از الگوریتم ژنتیک و الگوریتم جانسون برای تولید بی‌زائده‌ها استفاده شده و دو حالت *full-reducts* و *object-related-reducts* انتخاب گردیده است.

- **تولید قوانین**

تولید بی‌زائده‌ها برای استخراج قوانین تصمیم استفاده می‌شوند. در سمت چپ قوانین تصمیم، ترکیبی از ارزش‌های شاخص‌هاست که معمولا شاخص تصمیم تمامی مشاهداتی که با این ترکیب شاخص‌ها همخوانی دارند برابر سمت راست قانون می‌باشد. قوانینی که از بی‌زائده‌ها استخراج می‌شوند می‌توانند برای کلاسه‌بندی داده‌ها بکار روند. مجموعه قوانین بدست آمده به عنوان یک ابزار کلاسه‌بندی بکار رفته و می‌توان از آن برای کلاسه-بندی داده‌های جدید استفاده نمود. کیفیت قوانین به بی‌زائده‌های بدست آمده بستگی دارد. سعی بر آنست که تا قوانینی که بخش بیشتری از مجموعه مرجع را پوشش می‌دهد تولید شوند. پوشش مجموعه مرجع با قوانین دارای عمومیت بیشتر منجر به کوچک شدن مجموعه قوانین می‌گردد.

- **ارزیابی قوانین**

معیارهای کیفی مربوط به قوانین تصمیم می‌تواند برای حذف برخی قوانین دارای کیفیت پایین به کار رود (آراهام و دیگران، ۲۰۰۹)[۴۸]. قوانین *IF-THEN* با خواندن ارزش هر یک از شاخص‌های موجود در بی‌زائده و ارتباط دادن آن به یک یا چند گروه تصمیم ساخته می‌شوند. قسمت دوم یک قانون یا «بخش آنگاه» یک تصمیم، تنها شامل یک گروه تصمیم می‌شود، مگر اینکه گروه تصمیم با توجه با شاخص‌های موجود در بی‌زائده، راف باشد. برای ارزیابی قوانین تولیدی از میزان عمومیت قانون یا *coverage* استفاده می‌شود که برابر درصد مشاهداتی است که ضمن تعلق به گروه تصمیم «بخش آنگاه» شرایط «بخش اگر» را نیز پوشش می‌دهد. همچنین استفاده از دقت قانون یا *accuracy* 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. تعریف دقیق این دو پارامتر ارزیابی در ادامه آمده است:

$$Coverage = \frac{\text{تعداد مشاهداتی که هم شرایط «بخش آنگاه» را دارند و هم شرایط «بخش اگر»}}{\text{تعداد کل مشاهدات دارای شرایط «بخش آنگاه»}}$$

$$Accuracy = \frac{\text{تعداد مشاهداتی که هم شرایط «بخش آنگاه» را دارند و هم شرایط «بخش اگر»}}{\text{تعداد کل مشاهدات دارای شرایط «بخش اگر»}}$$

- **انتخاب قوانین**

مجموعه قوانین را می‌توان براساس پارامترهای بالا فیلتر نمود. همچنین می‌توان پارامترهای آماری مربوط به هریک از قوانین را نیز استخراج کرده و آنها را با یکدیگر مقایسه نمود.

به منظور انتخاب قوانین، در این تحقیق ابتدا اعتبار سنجی موازی برای هر یک از مجموعه قوانین اجرا شده و نتایج با یکدیگر مقایسه شده است؛ سپس مجموعه قوانین هر یک از سیستم‌های تصمیم ، براساس شاخص‌های موجود فیلتر شده و دقت قوانین باقی مانده با استفاده از قابلیت *classify* محاسبه شده و بهترین قوانین انتخاب شده‌اند.

- **پیش‌بینی**

این مرحله آخرین مرحله رویکرد مدل‌سازی راف است. پس از آنکه براساس تکنیک‌های توضیح داده شده در بالا بهترین مدل با بالاترین دقت انتخاب گردید، در این مرحله یک الگوی کلاسه‌بندی و پیش‌بینی ارائه می‌شود. برای کلاسه‌بندی و پیش‌بینی کافی است بخش «اگر» قوانین تولید شده با مقادیر مربوط به شاخص‌های موقعیتی هر مشاهده مقایسه شده و سپس با مقایسه بخش «آنگاه» مناسب‌ترین ارزش مربوط به شاخص تصمیم به آن مشاهده تخصیص یابد.

در رابطه با استفاده از تئوری راف در مطالعات مالی، هانگ[۴۹] (۲۰۱۱) سیستم خبره[۵۰]ای برای انتخاب اتوماتیک سهام با استفاده از تکنیک درجه رابطه خاکستری (GRA)، روش خوشه بندی/ طبقه بندی -VRRS Index و تئوری مجموعه راف ارائه داد. نتیجه ارائه شده نشان می‌دهد که این سیستم خبره عملکرد بهتری نسبت به سایر سیستم‌های موجود دارد. تقوی و گل‌باز (۱۳۹۵) از ترکیب مدل‌های رگرسیون خطی، تئوری مجموعه راف، سیستم خاکستری و برنامه‌ریزی خطی برای پیش‌بینی روند آتی نسبت‌های مالی، انتخاب سهام، رتبه‌بندی شرکت‌ها و تخصیص سرمایه استفاده کرد. پرتفویی از سهام تشکیل و عملکرد آن با استفاده از معیار شارپ با شاخص کل و شاخص ۵۰ شرکت فعال‌تر مورد مقایسه قرار گرفت و نتایج نشان داد مدل ترکیبی پژوهش در دوره مورد بررسی، عملکرد کاراتری نسبت به شاخص‌کل و شاخص ۵۰ شرکت فعال‌تر داشته است. سارنج و دیگران (۱۳۹۶) روشی مبتنی بر تئوری مجموعه‌های راف و با استفاده از شاخص‌های تحلیل تکنیکی جهت پیش‌بینی قیمت سهام ارائه نموده‌اند. محققان با استفاده از تئوری مجموعه‌های راف و ترکیب روش‌های مختلف گسسته‌سازی داده‌ها و تولید بی زائده را بر اساس داده‌های یادگیری، قواعد پیش‌بینی استخراج و قدرت پیش‌بینی روش‌های مختلف بر اساس داده‌های کنترل محاسبه کردند. 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 کرد.

۴- روش شناسی

۴-۱- داده‌ها و قلمرو تحقیق

تحقیق حاضر از لحاظ هدف تحقیقی بنیادی است چرا که در این تحقیق محقق با ارائه مدلی جدید در حوزه ارزیابی صندوق‌های سرمایه‌گذاری بدنبال جابجا کردن مرزهای دانش در این حوزه می‌باشد. از سوی دیگر این پژوهش یک تحقیق کاربردی است؛ به این دلیل که الگوی پیشنهادی به صورت اجرایی و عملی در حوزه تصمیم‌گیری در خصوص انتخاب صندوق سرمایه‌گذاری قابل استفاده خواهد بود. جهت تعیین وضعیت صندوق-های سرمایه‌گذاری در آینده و ایجاد یک سیستم خبره تصمیم‌گیری در این حوزه از روش تحقیق ترکیبی تشریحی استفاده خواهد شد. از داده‌ها، مستندات، مدارک، گزارش‌ها و مکتوبات موجود در سازمان بورس اوراق بهادار، نهادهای وابسته به آن و اطلاعات منتشره توسط صندوق‌های سرمایه‌گذاری به همراه نتایج حاصل از مطالعات قبلی و نیز پرسشنامه جهت گردآوری داده ها استفاده می‌شود. صندوق‌های سرمایه‌گذاری از سال ۱۳۸۶ در ایران شروع به فعالیت کرده‌اند، اما اطلاعات برخی از صندوق‌ها از سال ۱۳۹۰ به بعد در اختیار می-باشد و همچنین بسیاری از صندوق‌ها از سال ۱۳۹۰ به بعد تشکیل شدند، لذا قلمرو زمانی تحقیق از سال ۱۳۹۰ تا ۱۳۹۶ می‌باشد. داده‌ها از بین صندوق‌های سرمایه‌گذاری فعال سهام که دارای مجوز تأسیس و فعالیت از سازمان بورس و اوراق بهادار و تحت نظارت آن سازمان در کشور جمهوری اسلامی ایران گردآوری می‌شود. بیشترین صندوق‌های سرمایه‌گذاری در ایران از نوع صندوق‌های سرمایه‌گذاری مشترک هستند. از بین ۲۱۴ صندوق سرمایه‌گذاری مشترک، ۱۲۵ صندوق سرمایه‌گذاری در سهام، ۷۳ صندوق سرمایه‌گذاری با درآمد ثابت و ۱۶ صندوق سرمایه‌گذاری مختلط وجود دارد. در این بررسی صندوق‌های سرمایه‌گذاری در سهام به عنوان جامعه آماری انتخاب شده است. در ایران، ۱۲۵ صندوق سرمایه‌گذاری در سهام وجود دارد که ۵۸ درصد کل صندوق-های سرمایه‌گذاری در ایران را تشکیل می‌دهند؛ همچنین ۷۴ صندوق سرمایه‌گذاری که اطلاعات آنها موجود بود، انتخاب و مورد بررسی قرار گرفت. از دلایل انتخاب صندوق‌های سرمایه‌گذاری مشترک در سهام به عنوان جامعه آماری این می‌باشد که این نوع صندوق‌ها بیشترین تعداد صندوق‌های سـرمایه‌گذاری در ایران را تشکیل می‌دهند و از سوی دیگر ریسک این نوع صندوق‌ها نسبت به سایر صندوق‌ها بالاتر و به تبع بازدهی نسبتاً بالاتری نیز دارند که به هدف پژوهش نزدیک‌تر است. برای مطالعه ویژگی‌های موثر بر بازدهی و عملکرد لازم است تا همه ابعاد موثر بر بازدهی صندوق‌ها مورد توجه قرار گیرد تا بتوان یک سیستم خبره خوبی را طراحی نمود و به هدف تحقیق نزدیک شد.

۴-۲- نتایج حاصل از فرآیند مدل‌سازی راف در این مطالعه

گام اول: تعریف مساله

همانطور که بیان گردید، هدف این مطالعه استخراج موتور استنتاج سیستم خبره برای بررسی عملکرد صندوق-های سرمایه‌گذاری در سهام در ایران است. برای این منظور از تئوری راف و نرم افزار ROSETTA استفاده

گردید. بعد تعریف مساله باید مشخصه‌های اثرگذار بر عملکرد صندوق‌ها و سپس شاخص‌های عملکردی مشخص گردند.

گام دوم: انتخاب مشخصه‌ها اساسی و تکمیل داده‌ها

برای بررسی و استخراج مشخصه‌های موثر بر بازدهی و عملکرد صندوق‌های سرمایه‌گذاری، استفاده از مطالعات پیشین ضروری بود که برای این امر از اطلاعات موجود در کتب، مجلات، پایان‌نامه‌ها و سایر مستندات موجود در کتابخانه‌های تخصصی استفاده گردید. همچنین بر اساس موضوع تحقیق برخی داده‌ها و اطلاعات باید از اینترنت و پایگاه‌های داده‌ای مختلف استخراج می‌گردید که این مسئله بر اهمیت تحقیق و به روز بودن اطلاعات می‌افزاید. برخی اطلاعات مربوط به صندوق‌های سرمایه‌گذاری در سهام در واحد پردازش اطلاعات مالی ایران شرکت مدیریت فناوری بورس تهران (فیپیران) و سایت بازار بورس ایران موجود است که از این پایگاه‌ها استخراج شد. پس از شناسایی مشخصه‌های موثر بر عملکرد صندوق‌های سرمایه‌گذاری، به منظور کیفی کردن این مشخصه‌ها از نظر کارشناسان خبره در این زمینه بهره برداری شد. بر اساس نظر خبرگان، مشخصه‌های مشخص شده در پژوهش به صورت کیفی وارد مدل گردید. بر اساس اطلاعات موجود و استخراج شده از مرکز پردازش اطلاعات مالی ایران، مشخصه‌های مورد استفاده در این تحقیق بر اساس چگونگی محاسبه در جدول (۱) مشخص شده است.

جدول ۱: مشخصه‌های مورد بررسی در پژوهش

ردیف	عامل موثر بر بازدهی	نحوه اندازه‌گیری	نوع داده
۱	نسبت فعالیت معاملات	تقسیم حجم معاملات صندوق بر میانگین سرمایه صندوق	کمی
۲	نسبت فعالیت سرمایه	تقسیم یک دوم تفاضل ارزش صدور واحدهای سرمایه‌گذاری از ارزش ابطال واحدها بر متوسط ارزش صندوق	کمی
۳	درصد تملیک سرمایه‌گذاران حقیقی	مجموع درصد مالکیت سرمایه گذاران حقیقی	کمی
۴	تعداد صنعت و شرکت	تعداد صنعت‌هایی که صندوق در آنها سرمایه‌گذاری کرده	کمی
۵	نسبت سهام به کل دارایی صندوق	درصد سهام صندوق نسبت به کل وجه نقد، اوراق بهادار با درآمد ثابت، گواهی سپرده و سپرده بانکی و سایر دارایی ها	کمی
۶	اندازه صندوق (ارزش بازاری) و رتبه نقدشوندگی	میانگین ارزش صندوق در دوره مورد نظر	کمی
۷	عمر صندوق	تعداد ماه‌هایی که از تأسیس صندوق گذشته	کمی
۸	هزینه‌های کل	بر اساس اطلاعات هزینه‌های اعلام شده توسط هر صندوق	کمی
۹	هزینه صدور	بر اساس اطلاعات صندوق	کمی
۱۰	هزینه ابطال	بر اساس اطلاعات صندوق	کمی
۱۱	شارپ دوره	تقسیم بازدهی اضافی بر انحراف معیار	کمی

ردیف	عامل موثر بر بازدهی	نحوه اندازه‌گیری	نوع داده
۱۲	جنسن دوره	تقسیم بازدهی اضافی بر ریسک سیستماتیک	کمی
۱۳	ترینر دوره	تفاوت بین نرخ بازده صندوق و نرخ بازده بازار	کمی
۱۴	دوره برتر دوره	نسبت تعداد روزهایی که بازدهی صندوق بالاتر از بازده بازار است بر کل روزهای دوره	کمی

منبع: مرکز پردازش اطلاعات مالی ایران (فیپیران)

بنابراین مشخصه‌های موقعیتی این پژوهش، ۱۴ مشخصه مشخص شده در جدول (۱) می‌باشد. سپس در این مرحله داده‌ها، بر اساس روش‌های حذف داده‌های ناقص (*Remove*)، روش میانگین (*Mean*) و میانه شرطی (*Mean Conditional*) تکمیل گردیدند.

گام سوم: تخصیص کلاس‌های تصمیم

در این گام ابتدا برای اندازه‌گیری عملکرد و بازدهی صندوق‌های سرمایه‌گذاری، بر اساس نظر خبرگان، شاخص شارپ، جنسن و ترینر به عنوان شاخصهای تصمیم در نظر گرفته شد. این شاخص بر اساس نظر خبرگان به صورت کیفی طیف‌بندی گردید و به صورت بسیار عالی، عالی، متوسط، ضعیف و بسیار ضعیف مورد بررسی قرار گرفت. این طبقه‌بندی طیفی به این صورت انجام شده است که مقادیر شاخص‌های عملکردی از حداکثری به حداقلی طبقه‌بندی گردید و به ۵ دسته تقسیم‌بندی گردید و به صورت طیف‌های بسیار خوب، خوب، متوسط، ضعیف و خیلی ضعیف تقسیم‌بندی گردید. از سوی دیگر، بر اساس نظر خبرگان، بازده سرمایه‌گذاری بر اساس شاخص مورد ارزیابی و اطلاعات هر یک از صندوق‌ها در سال‌های مورد بررسی، به صورت بازده سرمایه‌گذاری یکساله مورد بررسی قرار داده می‌شود. جدول (۲) ویژگی شاخص اندازه‌گیری عملکرد صندوق‌های سرمایه‌گذاری و دامنه آن را نشان می‌دهد.

جدول ۲: شاخص ارزیابی عملکرد و بازده صندوق

ردیف	شاخص‌های عملکرد صندوق‌های سرمایه‌گذاری	تعریف عملیاتی	دامنه مقادیر به صورت کیفی	دامنه مقادیر به صورت کمی (گسسته‌سازی)
۱	نسبت شارپ	تقسیم بازدهی اضافی بر انحراف معیار	۱>شارپ>۱-	۵ طیف بسیار ضعیف تا بسیار خوب
۲	نسبت جنسن	تقسیم بازدهی اضافی بر ریسک سیستماتیک	۲۶۳۹۳/۶۲>ترینر>۲۳۷۰/۵۸-	۵ طیف بسیار ضعیف تا بسیار خوب
۳	نسبت ترینر	تفاوت بین نرخ بازده صندوق و نرخ بازده بازار	۲۶۴/۱۹>جنسن>۳۹/۸۶-	۵ طیف بسیار ضعیف تا بسیار خوب

منبع: یافته‌های محقق

در ادامه براساس مشخصه تصمیم معرفی شده، مدل‌سازی انجام شده و قواعد حاکم بر رفتار و ویژگی‌های تأثیرگذار بر عملکرد آتی صندوق‌های سرمایه‌گذاری برای انتخاب بهترین مدل استخراج گردید.

گام چهارم: فیلتر کردن داده‌ها

با توجه به اینکه داده‌های ورودی به نرم‌افزار به شکل فایل اکسل می‌باشند و ممکن است ستون‌های غیرضروری نظیر محاسبات و توضیحات نیز در آن قرار داشته باشند، لذا به کمک فیلترینگ می‌توان آنها را در محاسبات لحاظ نکرد. همچنین در برخی الگوریتم‌های تئوری راف که تنها با شاخص‌های از نوع عددی کار می‌کنند، می‌توان از این امکان برای در نظر نگرفتن موقت شاخص‌های از نوع متنی استفاده نمود.

گام پنجم: گسسته‌سازی داده‌ها

تئوری راف سعی دارد تا از داده‌های موجود در سیستم اطلاعاتی وزن و قدرت هریک از شاخص‌ها را در کلاسه‌بندی مشخص نماید. بنابراین این تئوری از برخی استراتژی‌ها برای گسسته‌نمودن ارزش‌های واقعی استفاده می‌کند که کیفیت الگوریتم‌های یادگیری به این استراتژی‌ها وابسته است. الگوریتم بولین، گسسته‌کردن دستی، الگوریتم آنتروپی، الگوریتم بازه‌های یکسان، الگوریتم ساده، الگوریتم شبه ساده، از جمله الگوریتم‌های گسسته‌سازی معرفی شده در قسمت قبل هستند که در این پژوهش از الگوریتم‌های بولین (*Boolean*)، انتروپی (*Entropy*)، یکسان (*Equal*) و نیو (*Nive*) برای گسسته‌کردن داده‌ها استفاده گردیده است.

گام ششم: انتخاب روش تقسیم‌بندی داده‌ها

در این گام، داده‌ها به دو گروه یادگیری و کنترل به منظور انجام مراحل بعدی تقسیم می‌شوند.

گام هفتم: استخراج بی‌زائده‌ها

بعد از مشخص نمودن جدول تصمیم، که حاوی داده‌های مربوط به مشخصه تصمیم می‌باشد، جدول تصمیم به عنوان ورودی نرم‌افزار *Rosetta* تعریف می‌شود. مراحل گسسته‌سازی برای تولید بی‌زائده‌ها به روش‌های مشخص شده و استخراج بی‌زائده‌ها توسط نرم افزار انجام می‌گیرد. از دو متد الگوریتم ژنتیک و جانسون برای محاسبه بی‌زائده‌ها بر اساس مشخصه های تصمیم شارپ،جنسن و ترینر بهره گرفته شد.

همانطور که در گام‌های قبلی عنوان گردید از سه الگوریتم برای تکمیل داده‌ها، چهار الگوریتم برای گسسته-سازی، دو الگوریتم برای استخراج بی‌زائده‌ها و یک الگوریتم برای تولید قانون استفاده گردید که با این حساب ۲۴ الگوی استخراج بی‌زائده استخراج گردید که در ادامه هر کدام از این الگوها توضیح داده می‌شود.

جدول ۳: استخراج بی‌زائده‌ها برای مدل بر اساس مشخصه تصمیم شارپ

شاخص تصمیم	الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی-زایده‌ها	بی‌زائده‌های حداقلی (به عنوان نمونه)
شارپ	Remove	Boolean	Genetic	۳۴	تعداد شرکت، هزینه ابطال، ترینر دوره
			Johnson	۱	درصد تملیک حقیقی، هزینه ابطال، ترینر دوره
		Entropy	Genetic	۳۹	اندازه صندوق، شارپ دوره
			Johnson	۱	اندازه صندوق، هزینه کل
		Equal	Genetic	۱۰۰	درصد تملیک حقیقی، درصد سهام به دارایی، اندازه صندوق، هزینه صدور، شارپ دوره، ترینر دوره، جنسن دوره
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه، درصد تملیک حقیقی، اندازه صندوق، هزینه صدور، شارپ دوره، جنسن دوره
		Nive	Genetic	۶۴	نسبت معاملاتی
			Johnson	۱	نسبت معاملاتی
	Mean	Boolean	Genetic	۵۴	نسبت سرمایه، هزینه کل، هزینه صدور، دوره برتر دوره
			Johnson	۱	نسبت سرمایه، نسبت سهام به دارایی، هزینه کل، شارپ دوره
		Entropy	Genetic	۴۵	تعداد شرکت، اندازه صندوق، عمر صندوق، هزینه ابطال
			Johnson	۱	تعداد شرکت، اندازه صندوق، هزینه کل
		Equal	Genetic	۸۹	درصد تملیک حقیقی، درصد سهام به دارایی، عمر صندوق، هزینه صدور، هزینه ابطال، شارپ دوره، ترینر دوره و دوره برتر دوره
			Johnson	۱	نسبت معاملاتی، درصد تملیک حقیقی، تعداد شرکت، اندازه صندوق، هزینه صدور، شارپ دوره، دوره برتر دوره
		Nive	Genetic	۶۹	نسبت معاملاتی، عمر صندوق
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
	Mean Conditional	Boolean	Genetic	۳۸	هزینه کل، شارپ دوره، جنسن دوره
			Johnson	۱	نسبت معاملاتی، هزینه کل، شارپ دوره
		Entropy	Genetic	۱۷	نسبت معاملاتی، هزینه صدور
			Johnson	۱	نسبت معاملاتی، اندازه صندوق

شاخص تصمیم	الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی-زایده‌ها	بی‌زائده‌های حداقلی (به عنوان نمونه)
		Equal	Genetic	۸۸	درصد تملیک حقیقی، درصد سهام به دارایی، عمر صندوق، هزینه کل، هزینه صدور، شارپ دوره، ترینر دوره، دوره برتر دوره
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه، تعداد شرکت، درصد سهام به دارایی، شارپ دوره، ترینر دوره، دوره برتر دوره
		Nive	Genetic	۷۵	نسبت معاملاتی، دوره برتر دوره
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه

منبع: یافته‌های تحقیق

جدول ۴: استخراج بی‌زائده‌ها برای مدل بر اساس مشخصه تصمیم جنسن

شاخص تصمیم	الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی (به عنوان نمونه)
جنسن	Remove	Boolean	Genetic	۳۲	اندازه صندوق، هزینه ابطال، جنسن دوره
			Johnson	۱	نسبت معاملاتی، هزینه ابطال، ترینر دوره
		Entropy	Genetic	۴۵	اندازه صندوق، هزینه کل
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
		Equal	Genetic	۶۸	نسبت معاملاتی، درصد تملیک حقیقی، تعداد شرکت، عمر صندوق، هزینه صدور، هزینه ابطال، شارپ دوره، دوره برتر دوره
			Johnson	۱	درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه ابطال، شارپ دوره، ترینر دوره، دوره برتر دوره
		Nive	Genetic	۷۶	نسبت معاملاتی، دوره برتر دوره
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
	Mean	Boolean	Genetic	۳۵	نسبت معاملاتی، درصد تملیک حقیقی، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه صدور، شارپ دوره، جنسن دوره، دوره

شاخص تصمیم	الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی (به عنوان نمونه)
					برتر دوره
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه، درصد تملک حقیقی، تعداد شرکت، اندازه صندوق، عمر صندوق، هزینه صدور، ترینر دوره، دوره برتر دوره
		Entropy	Genetic	۴۵	نسبت معاملاتی، هزینه صدور
			Johnson	۱	نسبت معاملاتی، هزینه صدور
		Equal	Genetic	۶۵	درصد تملیک حقیقی، تعداد شرکت، عمر صندوق، هزینه صدور، هزینه ابطال، شارپ دوره، ترینر دوره و دوره برتر دوره
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه، درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، شارپ دوره، ترینر دوره، دوره برتر دوره
		Nive	Genetic	۸۱	نسبت معاملاتی، تعداد شرکت
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
	Mean Conditional	Boolean	Genetic	۴۲	درصد تملیک حقیقی، هزینه ابطال، دوره برتر دوره
			Johnson	۱	نسبت معاملاتی، هزینه ابطال، ترینر دوره
		Entropy	Genetic	۴۱	نسبت سرمایه، اندازه صندوق
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
		Equal	Genetic	۷۸	نسبت سرمایه، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه کل، شارپ دوره، ترینر دوره، دوره برتر دوره
			Johnson	۱	نسبت سرمایه، درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه کل، شارپ دوره، دوره برتر دوره
		Nive	Genetic	۷۲	نسبت سرمایه، جنسن دوره
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه

منبع: یافته‌های تحقیق

جدول ۵: استخراج بی‌زائده‌ها برای مدل بر اساس مشخصه تصمیم ترینر

شاخص تصمیم	الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی (به عنوان نمونه)
ترینر	Remove	Boolean	Genetic	۳۶	نسبت معاملاتی، نسبت سرمایه، درصد تملیک حقیقی، تعداد شرکت، سهام به دارایی، اندازه صندوق، عمرصندوق، هزینه صدور، دوره برتر دوره
			Johnson	۱	نسبت معاملاتی، سهام به دارایی، عمر صندوق، هزینه کل، هزینه صدور، شارپ دوره، ترینر دوره، دوره برتر دوره
		Entropy	Genetic	۳۴	نسبت سرمایه، اندازه صندوق
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
		Equal	Genetic	۶۵	نسبت معاملاتی، درصد تملیک حقیقی، سهام به دارایی، هزینه کل، هزینه صدور، هزینه ابطال، شارپ دوره، دوره برتر دوره
			Johnson	۱	نسبت معاملاتی، تعداد شرکت، عمر صندوق، هزینه صدور، هزینه ابطال، شارپ دوره، ترینر دوره، دوره برتر دوره
		Nive	Genetic	۷۱	درصد تملیک حقیقی، نسبت سهام به دارایی
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
	Mean	Boolean	Genetic	۳۳	نسبت معاملاتی، نسبت سرمایه، درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه کل، هزینه صدور، هزینه ابطال
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه، درصد تملک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه کل، هزینه صدور، دوره برتر دوره
		Entropy	Genetic	۱۸	نسبت معاملاتی، نسبت سرمایه
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
		Equal	Genetic	۷۹	نسبت معاملاتی، درصد تملیک حقیقی، تعداد شرکت، سهام به دارایی، اندازه صندوق، هزینه صدور، شارپ دوره، جنسن دوره و دوره برتر دوره

شاخص تصمیم	الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی (به عنوان نمونه)
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه ابطال، جنسن دوره، دوره برتر دوره
		Nive	Genetic	۶۷	نسبت معاملاتی، هزینه کل
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
	Mean Conditional	Boolean	Genetic	۵۶	اندازه صندوق، هزینه ابطال، ترینر دوره
			Johnson	۱	نسبت معاملاتی، هزینه ابطال، ترینر دوره
		Entropy	Genetic	۲۸	نسبت سرمایه، اندازه صندوق
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه
		Equal	Genetic	۶۹	نسبت معاملاتی، درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، هزینه صدور، ترینر دوره، جنسن دوره، دوره برتر دوره
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، هزینه ابطال، شارپ دوره، جنسن دوره، دوره برتر دوره
		Nive	Genetic	۵۹	نسبت معاملاتی، اندازه صندوق
			Johnson	۱	نسبت معاملاتی، نسبت سرمایه

منبع: یافته‌های تحقیق

گام هفتم: استخراج قوانین و اعتبارسنجی قوانین

به منظور استخراج قوانین ابتدا مشاهدات به دو گروه مشاهدات یادگیری و گروه مشاهدات کنترل تقسیم-بندی گردید. در گروه مشاهدات یادگیری، قوانین بر اساس دو راهبرد *Full* و *ORR* و به روش الگوریتم ژنتیک و جانسون برای مدل مورد بررسی در این پژوهش استخراج شد که به تفکیک تعداد قوانین و نمونه‌ای از این قوانین استخراج شده، برای هر الگو مورد بررسی قرار می‌گیرد. قوانین با استفاده از بیشترین مقدار *RHS Coverage* فیلتر می‌شود. همچنین به منظور بررسی میزان دقت قوانین از گروه مشاهدات کنترل استفاده می‌گردد. هر چه شاخص دقت بالاتر باشد، قدرت پیش‌بینی قوانین حاصل از پردازش داده‌های یادگیری بالاتر می‌باشد. بدین منظور درصد اعتبار قانون مشخص می‌شود. بر اساس یافته‌ها، جدول (۶) بی‌زائده‌های استخراج شده را برای مدل با مشخصه تصمیم شارپ، جدول (۷) بی‌زائده‌های استخراج شده را برای مدل با مشخصه تصمیم جنسن و جدول

(۸) بی‌زائده‌های استخراج شده را برای مدل با مشخصه تصمیم ترینر، بر اساس الگوهای مختلف، به همراه تعداد قانون‌ها و همچنین درصد اعتبار آنها از بیشترین به کمترین درصد نشان می‌دهد.

زمانی که شاخص تصمیم آتی صندوق سرمایه‌گذاری شارپ باشد، بیشترین دقت پیش‌بینی بر اساس الگوی استخراج بی‌زائده‌ها، مربوط به مدل *Remove/ Nive/ Genetic/ Full* می‌باشد.

زمانی که شاخص تصمیم آتی صندوق سرمایه‌گذاری جنسن باشد، بر اساس الگوی استخراج بی‌زائده‌ها، بیشترین دقت پیش‌بینی مـربوط به مدل *Mean Conditional /Nive/Genetic/Full* می‌باشد.

و زمانی که شاخص تصمیم آتی صندوق سرمایه‌گذاری ترینر باشد، بر اساس الگوی بی‌زائده‌ها، بیشترین دقت پیش‌بینی مربوط به مدل Remove / Nive / Genetic/ Full می‌باشد.

جدول ۶: مشخصه‌های موقعیتی، تعداد قانون و درصد اعتبار در مدل با مشخصه تصمیم شارپ

الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی‌زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی(به عنوان نمونه)	تعداد قانون	اعتبار قانون
Remove	Nive	Genetic	۵۴	نسبت سرمایه، هزینه کل، هزینه صدور، دوره برتر دوره	۸۱۴۲	**۸۰**
Mean Conditional	Nive	Genetic	۷۵	نسبت معاملاتی، دوره برتر دوره	۶۵۶۴	**۷۹/۵۶**
Mean	Nive	Genetic	۶۹	نسبت معاملاتی، عمر صندوق	۱۲۹۳۰	**۷۵/۲۶**
Mean Conditional	Entropy	Genetic	۱۷	نسبت معاملاتی، هزینه صدور	۶۵۸۱	**۶۷/۷۴**
Remove	Equal	Genetic	۱۰۰	درصد تملیک حقیقی، درصد سهام به دارایی، اندازه صندوق، هزینه صدور، شارپ دوره، ترینر دوره، جنسن دوره	۱۹۷۸۹	**۶۶/۲۵**
Remove	Entropy	Genetic	۳۹	اندازه صندوق، شارپ دوره	۵۳۲۶	**۶۳/۷۵**
Mean	Equal	Genetic	۸۹	درصد تملیک حقیقی، درصد سهام به دارایی، عمر صندوق، هزینه صدور، هزینه ابطال، شارپ دوره، ترینر دوره و دوره برتر دوره	۱۸۸۱۲	**۶۲/۳۶**
Remove	Nive	Johnson	۱	نسبت سرمایه، نسبت سهام	۱۰۲	**۵۷/۵**

الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی‌زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی(به عنوان نمونه)	تعداد قانون	اعتبار قانون
				به دارایی، هزینه کل، شارپ دوره		
Mean	Entropy	Genetic	۴۵	تعداد شرکت، اندازه صندوق، عمر صندوق، هزینه ابطال	۷۴۲۹	**۵۶/۹۸**
Mean Conditional	Nive	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۲۵	**۵۵/۹۱**
Mean Conditional	Equal	Genetic	۸۸	درصد تملیک حقیقی، درصد سهام به دارایی، هزینه کل، هزینه صدور، شارپ دوره، ترینر دوره، دوره برتر دوره	۲۰۵۴۷	**۵۳/۷۶**
Remove	Equal	Johnson	۱	نسبت معاملاتی، نسبت سرمایه، درصد تملیک حقیقی، اندازه صندوق، هزینه صدور، شارپ دوره، جنسن دوره	۱۳۶	**۳۵/۷۵**
Remove	Boolean	Genetic	۳۴	تعداد شرکت، هزینه ابطال، ترینر دوره	۶۹۴۱	**۲۶**
Mean Conditional	Equal	Johnson	۱	نسبت معاملاتی، نسبت سرمایه، تعداد شرکت، درصد سهام به دارایی، شارپ دوره، ترینر دوره، دوره برتر دوره	۱۸۰	**۱۸/۲۷**
Mean	Equal	Johnson	۱	نسبت معاملاتی، درصد تملیک حقیقی، تعداد شرکت، اندازه صندوق، هزینه صدور، شارپ دوره، دوره برتر دوره	۱۷۷	**۷/۵۲**
Mean Conditional	Entropy	Johnson	۱	نسبت معاملاتی، اندازه صندوق	۱۷۴	**۵/۳۷**
Remove	Entropy	Johnson	۱	اندازه صندوق، هزینه کل	۱۵۶	**۵**
Mean Conditional	Boolean	Genetic	۳۸	هزینه کل، شارپ دوره، جنسن دوره	۲۰۳۱	**۳/۲۲**

الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی‌زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی(به عنوان نمونه)	تعداد قانون	اعتبار قانون
Mean	Boolean	Genetic	۶۴	نسبت معاملاتی	۳۵۴۲	**۲/۱۵**
Mean Conditional	Boolean	Johnson	۱	نسبت معاملاتی، هزینه کل، شارپ دوره	۱۸۴	**۲/۱۵**
Mean	Boolean	Johnson	۱	نسبت معاملاتی	۱۸۵	**۱/۰۷**
Mean	Entropy	Johnson	۱	تعداد شرکت، اندازه صندوق، هزینه کل	۱۸۴	**۱/۰۷**
Remove	Boolean	Johnson	۱	درصد تملیک حقیقی، هزینه ابطال، ترینر دوره	۱۶۲	۰
Mean	Nive	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۸۶	۰

منبع: یافته‌های تحقیق

جدول ۷: مشخصه‌های موقعیتی، تعداد قانون و درصد اعتبار در مدل با مشخصه تصمیم جنسن

الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی(به عنوان نمونه)	تعداد قانون	اعتبار قانون
Mean Conditional	Nive	Genetic	۷۲	نسبت سرمایه، جنسن دوره	۱۱۶۴۶	**۶۲/۳۶**
Remove	Nive	Genetic	۷۶	نسبت معاملاتی، دوره برتر دوره	۷۸۶۹	**۵۸/۰۲**
Mean	Nive	Genetic	۸۱	نسبت معاملاتی، تعداد شرکت	۱۲۶۵۴	**۵۵/۹۱**
Remove	Nive	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۲۱	**۳۴/۵۶**
Mean Conditional	Entropy	Genetic	۴۱	نسبت سرمایه، اندازه صندوق	۶۸۱۵	**۳۲/۲۵**
Remove	Equal	Genetic	۶۸	نسبت معاملاتی، درصد تملیک حقیقی، تعداد شرکت، عمر صندوق، هزینه صدور، هزینه ابطال، شارپ دوره، دوره برتر دوره	۱۵۷۳۰	**۳۰/۸۶**
Mean	Boolean	Genetic	۳۵	نسبت معاملاتی، درصد	۱۲۶۹۳	**۲۷/۹۵**

الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی(به عنوان نمونه)	تعداد قانون	اعتبار قانون
				تملیک حقیقی، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه صدور، شارپ دوره، جنسن دوره، دوره برتر دوره		
Mean	Entropy	Genetic	۴۵	نسبت معاملاتی، هزینه صدور	۸۳۷۷	**۲۳/۶۵**
Remove	Entropy	Genetic	۴۵	اندازه صندوق، هزینه کل	۵۱۵۶	**۲۳/۴۵**
Mean	Equal	Genetic	۶۵	درصد تملیک حقیقی، تعداد شرکت، عمر صندوق، هزینه صدور، هزینه ابطال، شارپ دوره، ترینر دوره و دوره برتر دوره	۲۰۵۲۱	**۲۱/۵**
Mean Conditional	Equal	Genetic	۷۸	نسبت سرمایه، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه کل، شارپ دوره، ترینر دوره، دوره برتر دوره	۱۴۶۶۴	**۲۰/۴۳**
Remove	Equal	Johnson	۱	درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه ابطال، شارپ دوره، ترینر دوره، دوره برتر دوره	۱۶۱	**۴/۹۳**
Mean Conditional	Boolean	Genetic	۴۲	درصد تملیک حقیقی، هزینه ابطال، دوره برتر دوره	۱۰۹۳۹	**۳/۲۲**
Remove	Entropy	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۶۰	**۲/۴۶**
Mean	Boolean	Johnson	۱	نسبت معاملاتی، نسبت سرمایه، درصد تملک حقیقی، تعداد شرکت، اندازه	۱۸۷	**۲/۱۵**

الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی(به عنوان نمونه)	تعداد قانون	اعتبار قانون
				صندوق، عمر صندوق، هزینه صدور، ترینر دوره، دوره برتر دوره		
Mean	Equal	Johnson	۱	نسبت معاملاتی، نسبت سرمایه، درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، شارپ دوره، ترینر دوره، دوره برتر دوره	۱۸۵	**۲/۱۵**
Mean Conditional	Equal	Johnson	۱	نسبت سرمایه، درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه کل، شارپ دوره، دوره برتر دوره	۱۸۶	**۲/۱۵**
Remove	Boolean	Genetic	۳۲	اندازه صندوق، هزینه ابطال، جنسن دوره	۵۸۳۳	**۱/۲۳**
Mean	Entropy	Johnson	۱	نسبت معاملاتی، هزینه صدور	۱۸۷	**۱/۰۷**
Mean	Nive	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۸۷	**۱/۰۷**
Mean Conditional	Boolean	Johnson	۱	نسبت معاملاتی، هزینه ابطال، ترینر دوره	۱۸۶	**۱/۰۷**
Remove	Boolean	Johnson	۱	نسبت معاملاتی، هزینه ابطال، ترینر دوره	۱۶۲	۰
Mean Conditional	Entropy	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۸۶	۰
Mean Conditional	Nive	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۸۶	۰

منبع: یافته‌های تحقیق

جدول ۸: مشخصه‌های موقعیتی، تعداد قانون و درصد اعتبار در مدل با مشخصه تصمیم ترینر

الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی(به عنوان نمونه)	تعداد قانون	اعتبار قانون
Remove	Nive	Genetic	۷۱	درصد تملیک حقیقی، نسبت سهام به دارایی	۹۳۱۹	**۶۵/۴۳**
Mean	Nive	Genetic	۶۷	نسبت معاملاتی، هزینه کل	۱۲۲۷۹	**۵۴/۸۳**
Mean Conditional	Nive	Genetic	۵۹	نسبت معاملاتی، اندازه صندوق	۱۱۳۲۷	**۵۱/۶۱**
Remove	Equal	Genetic	۶۵	نسبت معاملاتی، درصد تملیک حقیقی، سهام به دارایی، هزینه کل، هزینه صدور، هزینه ابطال، شارپ دوره، دوره برتر دوره	۱۷۰۵۹	**۳۵/۸**
Mean Conditional	Entropy	Genetic	۲۸	نسبت سرمایه، اندازه صندوق	۵۱۸۶	**۳۴/۴**
Remove	Nive	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۲۵	**۳۳/۳۳**
Mean Conditional	Equal	Genetic	۶۹	نسبت معاملاتی، درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، هزینه صدور، ترینر دوره، جنسن دوره، دوره برتر دوره	۲۲۶۱۸	**۳۱/۱۸**
Remove	Boolean	Genetic	۳۶	نسبت معاملاتی، نسبت سرمایه، درصد تملیک حقیقی، تعداد شرکت، سهام به دارایی، اندازه صندوق، عمرصندوق، هزینه صدور، دوره برتر دوره	۶۹۱۵	**۲۵/۹۲**
Mean	Equal	Genetic	۷۹	نسبت معاملاتی، درصد تملیک حقیقی، تعداد شرکت، سهام به دارایی،	۲۱۸۵۸	**۲۵/۸**

الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی(به عنوان نمونه)	تعداد قانون	اعتبار قانون
				اندازه صندوق، هزینه صدور، شارپ دوره، جنسن دوره و دوره برتر دوره		
Remove	Entropy	Genetic	۳۴	نسبت سرمایه، اندازه صندوق	۴۸۲۸	**۲۳/۴۵**
Mean	Entropy	Genetic	۱۸	نسبت معاملاتی، نسبت سرمایه	۴۹۸۹	**۲۰/۴۳**
Mean	Boolean	Genetic	۳۳	نسبت معاملاتی، نسبت سرمایه، درصد تملیک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه کل، هزینه صدور، هزینه ابطال	۱۰۷۱۸	**۱۷/۲۰**
Remove	Entropy	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۶۲	**۶/۱۷**
Remove	Boolean	Johnson	۱	نسبت معاملاتی، سهام به دارایی، عمر صندوق، هزینه کل، هزینه صدور، شارپ دوره، ترینر دوره، دوره برتر دوره	۱۵۵	**۶/۱۷**
Remove	Equal	Johnson	۱	نسبت معاملاتی، تعداد شرکت، عمر صندوق، هزینه صدور، هزینه ابطال، شارپ دوره، ترینر دوره، دوره برتر دوره	۱۶۱	**۶/۱۷**
Mean	Boolean	Johnson	۱	نسبت معاملاتی، نسبت سرمایه، درصد تملک حقیقی، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه	۱۸۲	**۳/۲۲**

الگوریتم تکمیل داده‌ها	الگوریتم گسسته‌سازی	الگوریتم استخراج بی-زائده‌ها	تعداد بی‌زایده‌ها	بی‌زائده‌های حداقلی(به عنوان نمونه)	تعداد قانون	اعتبار قانون
				کل، هزینه صدور، دوره برتر دوره		
Mean	Equal	Johnson	۱	نسبت معاملاتی، نسبت سرمایه، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، عمر صندوق، هزینه ابطال، جنسن دوره، دوره برتر دوره	۱۸۵	**۳/۲۲**
Mean Conditional	Equal	Johnson	۱	نسبت معاملاتی، نسبت سرمایه، تعداد شرکت، درصد سهام به دارایی، اندازه صندوق، هزینه ابطال، شارپ دوره، جنسن دوره، دوره برتر دوره	۱۸۵	**۳/۲۲**
Mean	Entropy	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۸۷	**۲/۱۵**
Mean Conditional	Boolean	Genetic	۵۶	اندازه صندوق، هزینه ابطال، ترینر دوره	۸۹۳۲	**۲/۱۵**
Mean	Nive	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۸۷	**۱/۰۷**
Mean Conditional	Boolean	Johnson	۱	نسبت معاملاتی، هزینه ابطال، ترینر دوره	۱۸۵	**۱/۰۷**
Mean Conditional	Entropy	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۸۶	**۱/۰۷**
Mean Conditional	Nive	Johnson	۱	نسبت معاملاتی، نسبت سرمایه	۱۸۶	**۰**

منبع: یافته‌های تحقیق

۵- نتیجه‌گیری و پیشنهادات

صندوق‌های سرمایه‌گذاری به دلیل داشتن پتانسیل بهره‌برداری از رشد بازار سرمایه در واقع همبستگی بالایی با وضعیت بازار سرمایه و شاخصه‌های کلان اقتصادی دارند. در این میان تنوع صنایع موجود در بازار سرمایه و روندهای درآمدی مختلف این صنایع و سوق‌گیری صندوق‌ها به ایجاد پرتفوی متنوع سرمایه‌گذاری، موجب شده است تا این برآیند عملکرد صنایع موجود در صندوق‌ها متجلی باشد و همان‌گونه که از آمار و شواهد برمی‌آید، این صندوق‌ها به دلیل مدیریت متمرکز و ساختارهای رسمی خود طی سالیان اخیر همواره عملکرد بهتری از شاخص بازار و عموم سرمایه‌گذاران کسب کرده‌اند. توسعه این صندوق‌ها یکی از ملزومات کشیدن نقدینگی به بازار سرمایه است و شاخت صحیح این صندوق‌ها بخشی از این سازوکار هستند. این صندوق‌های سرمایه‌گذاری با ایفای نقش واسطه مالی، سرمایه‌گذاری افراد غیرحرفه‌ای را از حالت مستقیم به غیرمستقیم تبدیل کرده و در این رهگذر مزایای متفاوتی را هم برای بازار سرمایه و هم برای سرمایه‌گذار فراهم می‌کنند. درمجموع صندوق‌های سرمایه‌گذاری با توجه به امکانات و توانایی‌هایی که در اختیار دارند، تلاش می‌کنند که سرمایه‌گذاری در بازار را جذاب‌تر کرده و فرصت‌ها و انتخاب‌های بیشتری را با ریسک کمتر و بازدهی بالاتر برای سرمایه‌گذاران فراهم کنند. در حال حاضر بیش از ۱۰۰ صندوق سرمایه‌گذاری در بازار سرمایه ایران فعال است که انتخاب این صندوق‌ها و رسیدن به سود و منفعت مناسب همراه با ریسک پایین‌تر مستلزم شناخت این صندوق‌ها بر اساس ویژگی‌های آنها است که افراد در انتخاب هر کدام از این صندوق‌ها باید به آن توجه داشته باشند.

این بررسی نیز با استفاده از مشخصه‌های تأثیرگذار بر بازدهی صندوق‌های سرمایه‌گذاری و شاخص تصمیم در خصوص بازدهی آتی صندوق‌ها به دنبال طراحی یک سیستم خبره با استفاده از تئوری راف بود. یک سیستم خبره، سیستمی باید باشد که تمام مشخصات و ویژگی‌های یک صندوق را در نظر بگیرد و به بررسی ارتباط همه این مشخصه‌ها بپردازد که در ادامه یک راهنمای خوب و مناسب برای تصمیم‌گیری سرمایه‌گذار در خصوص انتخاب صندوق‌ها باشد. همچنین بتواند بر اساس مشاهدات قبلی و نتایج به دست آمده از بررسی رابطه‌ها، به قواعدی برای پیش‌بینی رفتارهای آتی و اثرات مشخصه‌ها بر عملکرد آتی صندوق‌ها دست یابد. این سیستم کلیه عوامل را به طور همزمان در نظر می‌گیرد و از سوی دیگر با استفاده از مشاهداتی برای بررسی صحت نتایج، به تعیین درصد اعتبار قواعد به دست آمده می‌پردازد. در این مطالعه ویژگی‌ها یا مشخصه‌های تأثیرگذار بر عملکرد آتی صندوق‌های سرمایه‌گذاری تعیین گردید و بر اساس نتایج به دست آمده و با استفاده از تئوری راف یک سیستم خبره برای انتخاب صندوق سرمایه‌گذاری طراحی گردید. بر اساس نتایج به دست آمده بی‌زائده‌های هر الگو استخراج گردید و الگوهایی که بالاترین درصد اعتبار را داشتند معرفی گردیدند. بر اساس نتایج یک سیستم خبره طراحی شد که به عنوان پشتیبان تصمیم در انتخاب صندوق‌های سرمایه‌گذاری می‌تواند کمک‌کننده باشد. بنابراین در ادامه کار می‌توان «رابط کاربر» طراحی کرد که با وارد کردن مشخصات صندوق-های سرمایه‌گذاری، عملکرد آتی آنها را پیش‌بینی کرد. با توجه به نتایج به دست آمده از این پژوهش، در راستای بهبود عملکرد صندوق‌های سرمایه‌گذاری پیشنهادات کاربردی زیر ارائه می‌گردد: یکی از دغدغه‌های سرمایه-

گذاران، انتخاب صندوق سرمایه‌گذاری از بین تعداد زیاد صندوق‌های موجود می‌باشد؛ لذا سیستم‌های خبره می‌توانند به عنوان یک پشتیبان تصمیم در انتخاب صندوق‌ها کمک‌کننده باشند. از آنجا که موتور استنتاج، اصلی‌ترین قسمت هر سیستم خبره می‌باشد، لذا خروجی این تحقیق می‌تواند برای ساخت سیستم خبره انتخاب صندوق سرمایه‌گذاری بکار گرفته شود. با ساخت یک «رابط کاربری» ساده، می‌توان چنین سیستم خبره‌ای را تولید نمود و از آن به عنوان یک نرم‌افزار تجاری استفاده کرد. به بیان دیگر پس از ساخت رابط کاربری، سرمایه-گذار با واردکردن مشخصات صندوق سرمایه‌گذاری، می‌تواند میزان موفقیت صندوق را در یکسال آینده با توجه به شاخص‌های شارپ، جنسن و ترینر پیش‌بینی نماید. بنابراین مهم‌ترین پیشنهاد کاربردی حاصل از این تحقیق، استفاده از موتور استنتاج و کامل‌کردن سیستم خبره انتخاب صندوق سرمایه‌گذاری می‌باشد. همچنین پیشنهاد می‌شود در رابط کاربری امکان انتخاب مشخصه تصمیم توسط کاربر فراهم گردد تا کاربر براساس مشخصه تصمیم مدنظر خود، وضعیت صندوق در یکسال آینده را پیش‌بینی نموده و نتایج شاخص‌های مختلف را با یکدیگر مقایسه نماید. با توجه به اینکه این پژوهش در رابطه با صندوهای سرمایه‌گذاری در سهام انجام گرفته و اینکه مشخصه‌ها و ویژگی‌های موثر بر عملکرد سایر صندوق‌های سرمایه‌گذاری می‌تواند متفاوت باشد، لذا پژوهش مشابه در خصوص سایر صندوق‌های سرمایه‌گذاری می‌تواند زمینه مناسبی برای تحقیق و بررسی باشد. همچنین با توجه به اینکه مدل‌سازی در این پژوهش براساس تئوری مجموعه راف می‌باشد، پیشنهاد می‌شود با روش‌های دیگر مدل‌سازی انجام گیرد و نتایج با این پژوهش مقایسه گردد.

فهرست منابع

* تاج‌بر، علیرضا. (۱۳۹۴)، صندوق‌ها سرمایه‌گذاری، تهران: شرکت اطلاع رسانی و خدمات بورس، انتشارات بورس.
* کریمی، تورج.، صادقی‌مقدم، محمدرضا. (۱۳۹۳). مجموعه‌های راف و مجموعه‌های خاکستری: مبانی، کاربرد، نرم‌افزار، تهران: نشر: موسسه کتاب مهربان نشر، چاپ اول.
* تقوی، سید علی.، گل باز، لیلا. (۱۳۹۵). بررسی تأثیر توان اعتباری شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران بر نوسان‌پذیری ارزش بازار حقوق صاحبان سهام، فصلنامه راهبرد مدیریت مالی، دوره ۴، شماره ۱، ۱۴۱-۱۶۷.
* سارنج، علیرضا.، کریمی،تورج.، و شهرامی بابکان، مجید. (۱۳۹۶). کاربرد تئوری مجموعه‌های راف برای پیش‌بینی قیمت سهام (مطالعه موردی: بانک صادرات ایران)، راهبرد مدیریت مالی، شماره ۱۸، ۱۱۹- ۱۴۴.
* شینی زاده عمادی، شهرزاد. (۱۳۹۲). مطالعه رابطه بین نحوه مدیریت صندوق‌های سرمایه‌گذاری مشترک و عملکرد آنها، رساله دکتری، دانشکده اقتصاد دانشگاه مازندران.
* صادقی مقدم، محمدرضا.، علی بخشی، رضا.، خلیلی، الهام. (۱۳۹۵). ارزیابی صندوق‌های سرمایه گذاری مشترک برگزیده موجود در بازار سرمایه ی ایران با روشی ترکیبی از TOPSIS ، VIKOR و تکنیک مشابهت، فصلنامه تحقیقات مالی، دوره هفدهم - شماره ۲، ۲۵۹- ۲۸۲.

* Babalos, V., Mamatzakis, E.C., and Matousek, R. (2015). The performance of US equity mutual funds, Journal of Banking & Finance, vol. 52(C), pp. 217-229.
* Budiono D.P., and Martens M. (2009), Persistence in Mutual Fund Performance and Time-Varying Risk Exposures, Working Paper, SSRN.
* Evans, R., Fahlenbrach, R. (2012). Institutional Investors and Mutual Fund Governance: Evidence from Retail–Institutional Fund Twins, Review of Financial Studies, Volume 25, Issue 12, 3530-3571.
* Fang, Yi., and Wang, H. (2014). Fund Manager Characteristics and Performance, Investment Analysts Journal, MPRA Paper, No. 60013. Online at https://mpra.ub.uni-muenchen.de/60013/
* Ferreira, M.A., Keswani, A., Miguel, A.F., and S.B. Ramos, (2013). The Determinants of Mutual Fund Performance: A Cross-Country Study, Review of Finance, Vol. 17, Issue. 2, pp. 483–525,
* Hereil, P., Philippe, M., Moussavi, N., and Roncalli, T. (2010). Mutual Fund Ratings and Performance Persistence. Available at SSRN: https://ssrn.com/abstract=1749414 or http://dx.doi.org/10.2139/ssrn.1749414.
* Jabari R., Salehi Sedghiani J., Amiri M., Performance Evaluation and Portfolio Selection of Mutual Funds, Research in Operation and its Applications Journal, Vol. 9, pp. 1-19.
* Khalil, F., Ul Hassan, N., and Qamar. M.A. (2015). Managerial Attributes Effect on Mutual Fund Performance: Case from Pakistan, an Emerging Mutual Fund Market, European Online Journal of Natural and Social Sciences, Vol. 5, No. 2, pp. 432-441.
* Peters, H., Ferson, W., and Chen, Y. (2010). Measuring the timing ability and performance of bond mutual funds, Journal of Financial Economics, 98, 72-89.
* Qamar, H., Singh, S. (2016). Mutual fund performance prediction, Conference of IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER).
* Ruzgar, N.S., Ruzgar. B., and Unsal, F. (2015). Rough set theory and discriminant analysis to classify financial data, international Journal of Economics and statistics, Vol. 3, pp. 110-116.
* Trynor, J.L. (1965). How to Rate Management of Investment Funds, Harvard Business Review, XLIII, Vol. 43, pp. 63-7

یادداشت‌ها

[1] Rugh Set Theory
[2] Pierre Hereil
[3] Markov Modeling
[4] Helen Peters, Yong Chen and Wayne Ferson
[5] Budiono and Martens
[6] grey relational method
[7] Ferreira
[8] Yi Fang
[9] Babalos
[10] Stochastic Frontier Analysis
[11] Ruzgar
[12] Discriminant Analysis
[13] Mixed
[14] Rough Set Theory (RST)
[15] *Data reduct*
[16] Value set
[17] Indiscernibility
[18] Lower approximation and upper approximation
[19] Negative region

[20] Boundary
[21] Redundant
[22] Indispensable
[23] Core
[24] Reduct
[25] Significance of attributes
[26] Dispensable and indispensable
[27] An error of reduct approximation
[28] Condition classes
[29] Decision classes
[30] Supporting number
[31] Remove incompletes
[32] Mean/ Mode fill
[33] Conditional Mean/ Mode fill
[34] Combinatorial completion
[35] Conditional Combinatorial completion
[36] univariate and unsupervised
[37] univariate and supervised
[38] multivariate and supervised
[39] Boolean reasoning algorithm
[40] Manual discretization
[41] Entropy algorithm
[42] recursively partitioning
[43] Equal frequency binning
[44] Naive algorithm
[45] Semi naive algorithm
[46] support count
[47] Johnson's algorithm
[48] Abraham et al
[49] Haung
[50] Expert System