دانشکده مهندسی صنایع و مدیریت

گروه مدیریت

رشته : MBA

استخراج الگوهای نظام‌مند بازده سهام در بورس اوراق بهادار با استفاده از موج برگشتی نقدشوندگی

آزاده حجار

استاد راهنما:

دکتر رضا شیخ

استاد مشاور:

دکتر سید مجتبی میرلوحی

پایان‌نامه ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ۹۳

دانشگاه صنعتی شاهرود

دانشکده : مهندسی صنایع و مدیریت

گروه :MBA

پایان نامه کارشناسی ارشد خانم آزاده حجار

عنوان:

استخراج الگوهای نظام مند بازده سهام در بورس اوراق بهادار با استفاده از موج برگشتی نقدشوندگی

در تاریخ ۹۳/۱۱/۲۹ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مدیریت MBA مـورد ارزیابی و با درجه عالی مورد پذیرش قرار گرفت.

اساتید راهنما	امضاء	اساتید مشاور	امضاء
نام و نام خانوادگی :		نام و نام خانوادگی :	

اساتید داور	امضاء	نماینده تحصیلات تکمیلی	امضاء
نام و نام خانوادگی :		نام و نام خانوادگی :	
نام و نام خانوادگی :			

خدای را بسی شاکرم که از روی کرم، پدر و مادری فداکار نسیبم ساخته تا در سایه درخت پربار وجودشان بیاسایم و از ریشه آنها شاخ و برگ گیرم و از سایه وجودشان در راه کسب علم و دانش بهره جویم.

والدینی که بودنشان تاج افتخاری است بر سرم و نامشان دلیلی است بر بودنم، چرا که این دو وجود آموزگاران من در زندگی بودند.

تقدیم به وجود با ارزشتان...

تقدیرنامه

سپاس شایان نثار ایزد منان که توفیق را رفیق راهم ساخت تا این پایان نامه را به اتمام برسانم. از استاد فاضل و اندیشمند جناب آقای دکتر رضا شیخ به عنوان استاد راهنما که همواره اینجانب را مورد لطف و محبت خود قرار داده، کمال تشکر را دارم. هم چنین از جناب آقای دکتر سید مجتبی میرلوحی که با راهنمایی هایشان به این جانب در اتمام این پایان نامه یاری رساندند، سپاس گزارم.

تعهد نامه

اینجانب آزاده حجار دانشجوی دوره کارشناسی ارشد رشته .مدیریت MBA دانشکده مهندسی صنایع و مدیریت دانشگاه صنعتی شاهرود نویسنده پایان نامه استخراج الگوهای نظام‌مند بازده سهام در بورس اوراق بهادار با استفاده از موج برگشتی نقدشوندگی تحت راهنمائی دکتر رضا شیخ متعهد می‌شوم

تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .

در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .

مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .

کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .

حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.

در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .

در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است . تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .

استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده:

بورس اوراق بهادار نقش بسیار مهمی در توسعه سرمایه و تخصیص بهینه آن دارد و تلاش می‌کند که سرمایه‌های غیرکارا را به بخش‌های مولد اقتصادی هدایت کند. یکی از دغدغه‌های اصلی سرمایه‌گذاران بالفعل و بالقوه در بازار بورس اوراق بهادار کسب بازده مورد انتظار است، به همین علت همواره تحقیقات بسیاری در زمینه پیش‌بینی بازده سهام با استفاده از شاخص‌های مختلف اقتصادی صورت گرفته است. محققان مدل‌های بسیاری با رویکرد خطی، مثل مدل‌های تک‌عاملی و چندعاملی، طراحی نمودند اما به مرور زمان پی بردند که رفتار سهام آشوب‌گونه بوده و خطی نمی‌باشد، بنابراین استفاده از تکنیک‌های هوشمند و محاسبات نرم در پیش‌بینی رفتار سهام رونق گرفت. در این میان مهم‌ترین نکته برای سرمایه‌گذاران کشف الگوهای رفتاری بازده سهام می‌باشد، چرا که با استفاده از آن امکان تصمیم‌گیری در شرایط مطمئن‌تر، انتخاب زمان مناسب به منظور خرید و فروش سهام و ... مهیا می‌گردد. از این رو استفاده از تکنیک‌های داده‌کاوی می‌تواند کمک شایانی به فرآیند کشف الگوها و قوانین حاکم بر بازده سهام کند. در این پژوهش از تکنیک مجموعه راف استفاده شده است. این روش با بهره‌گیری از اطلاعات تاریخی به تجزیه و تحلیل رفتار بازده سهام پرداخته و بر اساس رفتار گذشته اقدام به پیش‌بینی آن می‌کند. علاوه بر آن با استفاده از یکی از معیارهای نوین سنجش نقدشوندگی به نام معیار عدم نقدشوندگی آمیهیود تعدیل شده و ترکیب تکنیک مجموعه راف با تئوری فازی، میزان نقدشوندگی سهام مورد بررسی قرار گرفته است.

فهرست مقالات مستخرج از پایان‌نامه

۱. ارزیابی نقدشوندگی سهام شرکت‌ها بر مبنای شاخص‌های کمی و عوامل مکنون با تکنیک بهینه‌سازی فازی - گروهی چند هدفه مبتنی بر سیستم نسبت. پذیرفته شده در مجله علمی-پژوهشی دانش سرمایه‌گذاری وابسته به انجمن مهندسی مالی ایران. چاپ در نشریه بهار ۹۴.

۲. استخراج الگوهای نظام‌مند بازده سهام در بورس اوراق بهادار با استفاده از موج برگشتی نقدشوندگی. در دست داوری.

۳. استخراج الگوهای نظام‌مند شاخص بورس اوراق بهادار با استفاده از موج برگشتی نقدشوندگی با استفاده از تکنیک مجموعه راف. در دست داوری

۴. ارزیابی نقدشوندگی سهام شرکت‌ها بر مبنای شاخص‌های کمی و عوامل مکنون با استفاده از تکنیک بهینه‌سازی مولتی مورا مبتنی بر فازی 2-tuple. در دست داوری.

۵. انتخاب سهام بر مبنای ریسک سرمایه‌گذاران با توجه به شاخص‌های نقدشوندگی. در دست داوری.

۶. تجزیه و تحلیل رفتار بازده سهام در بورس اوراق بهادار با استفاده از تکنیک مجموعه راف تحت سلطه. در دست داوری.

۷. رتبه‌بندی نقدشوندگی سهام شرکت‌ها با رویکرد مدل توافقی اصلاح شده. در دست داوری.

فهرست مطالب

فصل اول: کلیات تحقیق ---------- ۱

۱-۱ مقدمه ---------- ۳

۱-۲ تشریح و بیان موضوع تحقیق ---------- ۴

۱-۳ اهمیت و ضرورت تحقیق ---------- ۶

۱-۴ سؤالات تحقیق ---------- ۷

۱-۵ اهداف تحقیق ---------- ۸

۱-۶ نوع تحقیق ---------- ۸

۱-۷ روش تحقیق ---------- ۹

۱-۸ قلمرو تحقیق ---------- ۱۰

۱-۹ نوآوری تحقیق ---------- ۱۰

۱-۱۰ تعریف مختصر اصطلاحات، مفاهیم و واژه‌های تخصصی ---------- ۱۲

فصل دوم: مبانی نظری و ادبیات تحقیق ---------- ۱۵

۲-۱ بخش اول: مفاهیم بنیادی ---------- ۱۷

۲-۱-۱ مقدمه ---------- ۱۷

۲-۱-۲ بازارهای مالی و طبقه‌بندی آن‌ها ---------- ۱۷

۲-۱-۳ تعریف سهام و انواع آن ---------- ۱۹

۲-۱-۴ بازده سهام ---------- ۲۱

۲-۱-۵ نقدشوندگی سهام ---------- ۲۳

۲-۱-۶ بازار کارا و امکان پیش‌بینی قیمت‌ها ---------- ۲۸

۲-۱-۷ نظریه آشوب ---------- ۳۶

۲-۲ بخش دوم: انواع روش‌های پیش‌بینی ---- ۴۰
۲-۲-۱ روش‌های کلاسیک پیش‌بینی ---- ۴۰
۲-۲-۱-۱ رویکرد تجزیه و تحلیل بنیادین ---- ۴۰
۲-۲-۱-۲ رویکرد تجزیه و تحلیل تکنیکال ---- ۴۱
۲-۲-۱-۳ رویکرد مبتنی بر نظریه‌های مدرن مالی ---- ۴۲
۲-۲-۱-۳-۱ مدل قیمت‌گذاری دارایی‌های سرمایه‌ایی ---- ۴۳
۲-۲-۱-۳-۲ مدل فاما و فرنچ ---- ۴۷
۲-۲-۱-۳-۳ سایر مدل‌های چندعاملی ---- ۴۹
۲-۲-۲ روش‌های مدرن پیش‌بینی ---- ۵۳
۲-۲-۲-۱ رویکرد بی‌نظمی و پویایی غیرخطی ---- ۵۴
۲-۲-۳ پیشینه تحقیق ---- ۵۵
۲-۲-۳-۱ پژوهش‌های انجام گرفته در ایران ---- ۵۶
۲-۲-۳-۲ پژوهش‌های انجام گرفته در خارج از کشور ---- ۵۹
فصل سوم: متدولوژی تحقیق ---- ۶۵
۳-۱ مروری اجمالی بر داده‌کاوی ---- ۶۷
۳-۱-۱ مقدمه ---- ۶۷
۳-۱-۲ پیشینه داده‌کاوی ---- ۶۸
۳-۱-۳ مفاهیم کلی داده‌کاوی ---- ۷۰
۳-۱-۴ ابزارهای داده‌کاوی ---- ۷۵
۳-۱-۵ انواع روش‌های داده‌کاوی ---- ۷۶
۳-۱-۶ دلایل گزینش تئوری مجموعه راف ---- ۸۰

۳-۲ تئوری مجموعه راف ---- ۸۱

۳-۲-۱ سیستم‌های اطلاعاتی و رابطه غیرقابل متمایزسازی ---- ۸۴

۳-۲-۲ تقریب بالایی و تقریب پایینی ---- ۸۵

۳-۲-۳ وابستگی صفات، دقت و کیفیت تقریب ---- ۸۷

۳-۲-۴ محاسبه تقلیل‌ها و هسته با استفاده از ماتریس متمایزسازی ---- ۹۰

۳-۲-۵ جدول تصمیم و استخراج قوانین ---- ۹۳

۳-۲-۶ ارزیابی قواعد ---- ۹۵

فصل چهارم: تجزیه و تحلیل داده‌ها ---- ۹۹

۴-۱ مقدمه ---- ۱۰۱

۴-۲ جامعه‌آماری ---- ۱۰۱

۴-۳ متغیرهای تحقیق ---- ۱۰۳

۴-۳-۱ توصیف متغیرهای تحقیق ---- ۱۰۴

۴-۴ مطالعه موردی و تجزیه و تحلیل داده‌ها ---- ۱۰۹

۴-۴-۱ آماده‌سازی داده‌ها ---- ۱۱۰

۴-۴-۱-۱ کاهش مقادیر یک صفت خاصه ---- ۱۱۰

۴-۴-۲ اجرای تکنیک مجموعه راف ---- ۱۱۷

۴-۴-۲-۱ تشکیل سیستم اطلاعاتی ---- ۱۱۷

۴-۴-۲-۲ کلاس‌های هم‌ارزی و تقریب بالایی و پایینی ---- ۱۱۸

۴-۴-۲-۳ محاسبه تقلیل‌ها و استخراج قوانین (الگوریتم ژنتیک و 1Rهولت) ---- ۱۲۲

۴-۴-۲-۴ ارزیابی مدل با تکنیک K-Fold Cross-Validation ---- ۱۲۹

۴-۵ تفسیر قوانین و نتایج حاصل از تحقیق ---- ۱۳۱

فصل پنجم : نتیجه‌گیری ---------- ۱۳۷

۵-۱ مقدمه ---------- ۱۳۹

۵-۲ نتیجه‌گیری ---------- ۱۳۹

۵-۳ محدودیت‌های تحقیق ---------- ۱۴۲

۵-۴ پیشنهادات ---------- ۱۴۲

۵-۴-۱ پیشنهادات کاربردی ---------- ۱۴۲

۵-۴-۲ پیشنهادات برای تحقیقات آتی ---------- ۱۴۲

منابع ---------- ۱۴۵

فهرست جداول و اشکال

شکل شماره (۳-۱) جایگاه داده‌کاوی در فرآیند کشف دانش ---------- ۷۲

شکل شماره (۳-۲) مراحل فرآیند داده‌کاوی ---------- ۷۳

شکل شماره (۳-۳) طبقه‌بندی انواع روش‌های داده‌کاوی ---------- ۷۷

جدول (۳-۱) جدول تصمیم در مجموعه راف ---------- ۹۴

جدول (۴-۱) حداقل و حداکثر ده شاخص مورد استفاده ---------- ۱۱۲

جدول (۴-۲) حداقل و حداکثر تغییرات ده شاخص مورد استفاده ---------- ۱۱۲

جدول (۴-۳) گسسته‌سازی داده‌های تغییرات ده شاخص با استفاده از روش استدلال بولی متعامد ۱۱۲

جدول (۴-۴) گسسته‌سازی داده‌های تغییرات ده شاخص با استفاده از روش آنتروپی ---------- ۱۱۴

جدول (۴-۵) گسسته‌سازی داده‌های تغییرات ده شاخص با استفاده از روش آنتروپی ---------- ۱۱۵

جدول (۴-۶) گسسته‌سازی داده‌های تغییرات ده شاخص با استفاده از روش آنتروپی ---------- ۱۱۶

جدول (۴-۷) گسسته‌سازی داده‌های تغییرات ده شاخص با روش گروه‌بندی فراوانی یکسان ---------- ۱۱۷

جدول (۴-۸) جدول تصمیم با متغیر تصمیم بازده سهام ---------- ۱۱۸

جدول (۴-۹) جدول تصمیم با متغیر تصمیم نقدشوندگی ---------- ۱۱۸

جدول (۴-۱۰) متغیرهای کلامی برای نقدشوندگی ---------- ۱۱۹

جدول (۴-۱۱) کلاس‌های هم‌ارزی، تقریب بالایی، تقریب پایینی، منطقه مرزی و منطقه منفی با متغیر تصمیم بازده سهام ---------- ۱۲۰

جدول (۴-۱۲) کلاس‌های هم‌ارزی، تقریب بالایی، تقریب پایینی، منطقه مرزی و منطقه منفی با متغیر تصمیم بازده سهام ---------- ۱۲۱

جدول (۴-۱۳) قوانین استخراجی و فاکتورهای پشتیبان، قطعیت و پوشش ---------- ۱۲۵

جدول (۴-۱۴) قوانین استخراجی و فاکتورهای پشتیبان، قطعیت و پوشش ---------- ۱۲۶

جدول (۴-۱۵) قوانین استخراجی و فاکتورهای پشتیبان، قطعیت و پوشش ---------- ۱۲۷

جدول (۴-۱۶) قوانین استخراجی و فاکتورهای پشتیبان، قطعیت و پوشش ---------- ۱۲۸

فصل اول

کلیات تحقیق

۱-۱ مقدمه :

تجهیز و تخصیص بهینه‌ی سرمایه، نقش بسیار مهمی در رشد و توسعه اقتصادی کشور دارد، خصوصاً در کشورهای در حال توسعه؛ چرا که با تهدیدات عدیده‌ای روبه‌رو هستند و به منظور حل مشکلات اقتصادی خود، نیازمند راهکارهای مناسبی جهت استفاده بهتر از امکانات و ثروت‌های موجود می‌باشند؛ در این راستا، یکی از راهکارهای مناسب، بسط و توسعه سرمایه‌گذاری می‌باشد [۱].

بورس اوراق بهادار در تجهیز و تخصیص بهینه سرمایه و همچنین بسط و توسعه سرمایه‌گذاری جایگاه ویژه‌ای دارد و می‌تواند پس‌اندازها را به سرمایه‌گذاری در شرکت‌ها هدایت کند[۲]. توسعه سرمایه‌گذاری از یک‌سو موجب جذب سرمایه‌های غیر کارا و هدایت آن‌ها به بخش‌های مولد اقتصادی می‌گردد و از سوی دیگر با توجه به جهت‌گیری سرمایه‌گذاران (مبتنی بر ریسک و بازده)، سرمایه‌گذاری در صنایعی هدایت می‌شود که از سود بیشتر یا ریسک کمتری برخوردارند و این امر موجب تخصیص بهینه منابع می‌گردد[۳]. همان طور که گفته شد بازده سهام یکی از فاکتورهای بسیار مهم در جهت‌گیری سرمایه‌گذاران و انتخاب بهترین سرمایه‌گذاری است، بنابراین پیش‌بینی و مقایسه بازده سهام شرکت‌های مختلف یکی از روش‌های بهبود فرآیند سرمایه‌گذاری محسوب می‌شود. سهامداران برای تصمیم‌گیری در مورد خرید و فروش سهام شرکت‌ها نیازمند اطلاعاتی برای پیش‌بینی بازدهی سهام شرکت‌ها هستند[۴]. امروزه فرضیه قابلیت پیش‌بینی بازدهی سهام به عنوان واقعیت در مدیریت مالی مورد پذیرش قرار گرفته است [۵]، با این وجود بیشتر مطالعات در بازار سهام، بر قیمت سهام و رفتار آن در طول زمان متمرکز بوده است، اما به دلیل برخی ویژگی‌های تصادفی و ناخوشایند قیمت سهام، از قبیل متغیر بودن آن، اغلب پژوهش‌ها به جای قیمت بر بازده سهام تمرکز می‌کنند و در سال‌های اخیر وجود عوامل مؤثر بسیاری که به پیش‌بینی بازده سهام کمک می‌کنند مطرح شده است. بر اساس اطلاعات موجود در مورد شرکت، بازده سهام منعکس کننده انتظارات سرمایه‌گذاران از عملکرد آتی شرکت است. از آن‌جا که انگیزه و اولویت اصلی سرمایه‌گذاران کسب بازده مناسب می‌باشد، لذا با پیش‌بینی بازده سرمایه‌گذاری و ارائه

مدل‌هایی برای آن، می‌توان شرایط مطمئن‌تری در بازار سرمایه ایجاد کرد که به امر گسترش سرمایه‌گذاری در بازارهای مالی کمک می‌کند[۶].

۱-۲ تشریح و بیان موضوع تحقیق :

از آن‌جا که اولویت اصلی سرمایه‌گذاران بالقوه و بالفعل که بازیگران اصلی بازار بورس هستند، کسب بازده مورد انتظار است، یکی از معیارهای اساسی برای تصمیم‌گیری در امر سرمایه‌گذاری، بازده سهام می‌باشد. بازده سهام خود به تنهایی دارای محتوای اطلاعاتی بی‌شماری است و بیشتر سرمایه‌گذاران در تجزیه و تحلیل مالی و پیش‌بینی‌ها از آن استفاده می‌کنند. بازده سهام نه تنها معیار اصلی سرمایه‌گذاران در تصمیم‌گیری است، بلکه در برگیرنده اطلاعاتی نظیر ریسک، میزان نقدشوندگی، صرف نقدشوندگی و .... نیز می‌باشد. بنابراین ارزیابی و پیش‌بینی بازده بازار به عنوان ابزاری در راستای نیاز سرمایه‌گذاران به پوشش ریسک بازار می‌تواند کمک مؤثری در اخذ تصمیمات منطقی سرمایه‌گذاران به حساب آید[۷,۳]. بنابراین شناسایی عوامل تأثیرگزار بر بازده سهام همواره موضوع پژوهش‌های فراوانی در ادبیات مدیریت مالی و حسابداری بوده است و مدل‌های بسیاری در این زمینه ارائه شده است. با دانستن این نکته که یک یا چند عامل نرخ‌های بازدهی اوراق بهادار را تحت تأثیر قرار می‌دهند، هدف اصلی و اولیه تجزیه و تحلیل اوراق بهادار نسبت به تغییرات آنها است. به بیان رسمی به چنین رابطه‌ایی اصطلاحاً مدل عاملی نرخ‌های بازدهی اوراق بهادار گفته می‌شود [۸].

این مدل‌ها به دو گروه مدل‌های تک عاملی و مدل‌های چند عاملی تقسیم می‌شوند. مفهوم اساسی در مدل تک عاملی این است که تمامی اوراق بهادار از نوسانات عمومی بازار تأثیر می‌پذیرند، زیرا نیروهای اقتصادی مشابهی در آینده اکثر شرکت‌ها را تحت تأثیر قرار می‌دهد. یکی از معروف‌ترین این مدل‌ها مدل قیمت‌گذاری دارایی‌های سرمایه‌ایی (CAPM)[۱] می‌باشد. نتایج پژوهش‌های دو دهه‌ی گذشته در آمریکا،

[۱] : CAPM: Capital Asset Pricing Model

ژاپن و سایر کشورهای پیشرفته حاکی از آن است که این مدل توانایی‌های لازم برای پیش‌بینی بازده سهام را ندارد[۱۰,۹]. در واقع ریسک‌های بی‌شماری در ارتباط با شرکت وجود دارد، برخی از این ریسک‌ها شامل ریسک بازار، ریسک ورشکستگی و ریسک نقدینگی است، در حالی که در مدل CAPM تنها از یک عامل جهت تشریح مجموع ریسک‌ها استفاده می‌شود. از نظر منطقی، به نظر می‌رسد مدلی با عوامل گوناگون، ممکن است قدرت توصیف‌کنندگی و پیش‌بینی‌کنندگی بیشتری را فراهم آورد. به عبارت دیگر، عوامل اضافه، اجازه می‌دهند ریسک‌هایی توزیع شوند که شرکت با آن مواجه است. عامل ریسک بازار صرفاً ابعاد گوناگون ریسک را تجزیه می‌کند و قادر به بیان، تشریح و توصیف اثر انواع ریسک بر بازده نمی‌باشد. پس از آن فاما و فرنچ مدل سه عاملی خود را ارائه کردند که هدف اصلی در مدل‌های چند عاملی یافتن برخی از تأثیرات غیر بازاری است که منجر به حرکت توأم سهام با یکدیگر می‌شود چرا که با مطالعه آزمون‌های تجربی مدل تک عاملی CAPM ، مشاهده می‌شود که عوامل دیگری در توضیح بازده سهام نقش دارند. این مدل نیز در معرض انتقادهای فراوانی قرار گرفت که این انتقادها در چارچوب سه موضوع محوری داده‌کاوی، اشتباه انتخاب و روش برآورد ریسک سیستماتیک مطرح شد. پس از آن کارهارت[۲] با افزودن متغیر چهارم، عامل شتاب، مدل سه عاملی فاما و فرنچ را به مدلی چهار عاملی تبدیل کرد[۱۱,۹]. متأسفانه مدل‌های ذکر شده، همگی خطی هستند در حالی که بازار سهام سیستمی غیرخطی و آشوب‌گونه دارد که تحت تأثیر شرایط سیاسی، اقتصادی و روانشناسی است، لذا پیش‌بینی داده‌هایی که از این سیستم پیروی می‌کنند، نیازمند ابزارهای هوشمند و پیشرفته‌ایی هم‌چون محاسبات نرم[۳] و روش‌های ترکیبی است. تحقیقات اخیر نشان می‌دهد در طی چند دهه گذشته مدل‌های داده‌گرای محاسبات نرم حضور موفقی در مباحث مدیریت و مالی داشته‌اند و ایده آموزش برای حل مسائل شناسایی الگوهای پیچیده، هم‌چون بازار غیرخطی سهام با استفاده از دیدگاه عامل داده هوشمند برای محققان دانشگاهی بسیار چالش برانگیز و جذاب شده است. مدل‌های بسیاری به منظور پیش‌بینی بازده سهام با استفاده از محاسبات نرم قابل انعطاف غیرخطی، مانند شبکه‌های عصبی مصنوعی طراحی شده‌اند و

---

[۲] . Carhat’S Model
[۳] : Soft Computing

گرایش به افزایش دقت این مدل‌ها، محققان را بر آن می‌دارد که دست به ترکیب و تلفیق این روش‌ها بزنند. در این تحقیق با توجه به اینکه رفتار سری زمانی بازده روزانه سهام شرکت‌ها یک فرآیند تصادفی نیست، دارای حافظه می‌باشد و ماهیتی پیچیده و آشوب‌گونه دارد، سعی بر آن است که با سیستمی هوشمند و تلفیقی، تئوری راف ست، مدلی برای پیش‌بینی بازده سهام یکی از شرکت‌های حاضر در بورس اوراق بهادار ایران در جهت بهبود کفایت تصمیمات مالی، طراحی گردد[۱۲,۳].

تئوری راف ست یک ابزار قدرتمند ریاضی برای استدلال در موارد ابهام و نایقینی است که روش‌هایی را برای زدودن و کاستن اطلاعات و دانش نامربوط یا مازاد بر نیاز از پایگاه‌های داده مهیا می‌سازد. این فرآیند حذف داده‌های زائد، وظیفه اصلی سیستم بوده و بدون از دست دادن داده‌های اساسی پایگاه داده صورت می‌گیرد. در نتیجه تقلیل اطلاعات، مجموع‌هایی از قواعد تلخیص شده و پرمعنا حاصل می‌گردد که کار تصمیم‌گیرنده را بسیار ساده‌تر می‌کند. در حقیقت می‌توان گفت که مجموعه راف با کاهش فضای داده‌ها و برگزیدن ترم‌های مهم، یک نگاشت از فضای داده‌های خام به فضای سمانتیک (مفاهیم) انجام می‌دهد. لذا با توجه به رشد انفجاری حجم اطلاعات، مجموعه راف می‌تواند نقش بسیار مؤثری در سیستم‌های پشتیبان تصمیم‌گیری داشته باشد.

۱-۳ اهمیت و ضرورت تحقیق :

بدون تردید سرمایه‌گذاری در بورس، بخش مهمی از اقتصاد کشور را تشکیل می‌دهد و بدون شک بیشترین میزان سرمایه از طریق بازارهای سهام در سرتاسر جهان مبادله می‌شود و اقتصاد ملی به شدت متأثر از عملکرد بازار بورس است. هم‌چنین این بازار هم برای سرمایه‌گذاران حرفه‌ای و هم برای عموم مردم به عنوان یک ابزار سرمایه‌گذاری در دسترس است. امروزه با توجه به اهمیت و گسترش روز افزون بازارهای سرمایه در تجهیز و گردآوری سرمایه‌های کوچک فردی به سمت فعالیت‌های تولیدی، شناسایی رفتار سرمایه‌گذاران و متغیرهای تأثیرگذار بر قیمت و بازده سهام در این بازارها اهمیت زیادی پیدا کرده

است . نظر به این که بورس اوراق بهادار تهران، سازمانی در جهت تجهیز پس‌اندازها و هدایت آن به سمت سرمایه‌گذاری مولد و مفید به حال جامعه و اقتصاد کشور است، مطالعه پیرامون موضوع‌های وابسته به این سازمان از اهمیت ویژه‌ای برخوردار است. از آن جایی که بورس اوراق بهادار تهران در مدت زمان فعالیت خود با فراز و نشیب‌های فراوان دست به گریبان بوده است، بنابراین نیاز به تحقیقات و مطالعات گسترده در این زمینه ضروری به نظر می‌رسد. یکی از موضوعات مهم در ارتباط با سازمان بورس اوراق بهادار پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در این سازمان است.

بورس از پارامترهای کلان اقتصادی و غیراقتصادی بی‌شماری تأثیر می‌پذیرد، متعدد بودن عوامل مؤثر بر بازارهای سرمایه و ناشناخته بودن آن‌ها، موجب عدم اطمینان در زمینه سرمایه‌گذاری و پیش‌بینی بازده مورد انتظار شده است. نیاز به حضور مدلی غیرخطی و دقیق که توانایی پیش‌بینی بازده سهام را در شرایط عدم اطمینان و همچنین قدرت تأثیر دادن عوامل کیفی این فضای آشوب‌گونه را داشته باشد، به شدت احساس می‌شود. علاوه بر آن با توجه به روند رو به رشد اطلاعات در عصر کنونی، استفاده از روش‌های جدید داده‌کاوی و محاسبات نرم به منظور افزایش سرعت و قدرت پردازش اطلاعات امری ضروری و اجتناب‌ناپذیر است و به عنوان ابزاری قدرتمند و مفید در حوزه‌ی مدیریت مالی، توجه بسیاری از محققان را به خود جلب کرده است. در این پایان نامه سعی بر آن است که با استفاده از روش دقیق و پرکاربرد راف ست و تلفیق آن با سایر متدهای محاسبات نرم، از جمله سیستم فازی و .... مدل مناسبی به منظور پیش‌بینی بازده سهام بورس اوراق بهادار تهران طراحی گردد.

۱-۴ سؤالات تحقیق :

با توجه به الگوریتم تکنیک تئوری مجموعه راف دو دسته متغیر به نام‌های متغیرهای شرطی و متغیر تصمیم در این روش بکار گرفته می‌شوند. متغیرهای شرطی به منظور توصیف ویژگی‌ها و خصوصیات اشیا

استفاده می‌شوند و متغیر تصمیم به منظور بررسی و اخذ تصمیم مورد نظر. بنابراین در این پژوهش دو سوال زیر مطرح می‌شود:

۱. مهمترین متغیرهای شرطی و متغیر تصمیم در ارزیابی رفتار بازده سهام بورس اوراق بهادار تهران کدامند؟
۲. چگونه می‌توان با استفاده از تئوری راف ست قواعد منطقی را از رفتار بازده سهام بورس اوراق بهادار تهران استخراج نمود؟

۱-۵ اهداف تحقیق :

هدف اصلی از این تحقیق استخراج الگویی دقیق و کم خطا از پایگاه داده موجود در بورس اوراق بهادار تهران به منظور پیش‌بینی بازده سهام با استفاده از روش راف ست می‌باشد. در این میان شناسایی شاخص‌ها و عوامل مؤثر بر بازده سهام بورس اوراق بهادار نیز از دیگر اهداف این پژوهش محسوب می‌شود. در این تحقیق ارتباط میان بازده سهام و میزان نقدشوندگی آن، نحوه‌ی تأثیرگذاری این دو عامل بر یکدیگر و نقش نقدشوندگی سهام بر پیش‌بینی بازده آن نیز مورد بررسی قرار گرفته است.

۱- ۶ نوع تحقیق :

با توجه به ماهیت موضوع و هدف آن، تحقیق پیش‌رو از نوع تحقیقات کاربردی می‌باشد، زیرا نتایج حاصل از آن در تصمیم‌گیری‌های تخصیص بهینه مالی سرمایه‌گذاران نقش مهمی ایفا می‌کند. از منظر روش تحقیق، دارای چارچوبی توصیفی از نوع همبستگی است، زیرا به بررسی ارتباط میان متغیرهای شرطی و متغیر تصمیم می‌پردازد. نحوه‌ی گردآوری داده‌ها در این پژوهش به صورت کتابخانه‌ایی بوده و با استفاده

از اطلاعات سازمان بورس اوراق بهادار تهران، بانک مرکزی جمهوری اسلامی ایران، مدیریت اطلاعات انرژی آمریکا و شورای جهانی طلا جمع‌آوری شده‌اند.

۱-۷ روش تحقیق:

متدولوژی پژوهش حاضر را می‌توان به صورت زیر خلاصه نمود:

الف . مطالعه پیرامون ادبیات بازده سهام و سیر تاریخی آن، مدل‌های ارائه شده‌ی ابتدایی و معایب آن‌ها، چگونگی تکامل این مدل‌ها و رویکردهای نوین در این زمینه

ب . مطالعه در مورد نقدشوندگی سهام، انواع معیارهای سنجش نقدشوندگی، بررسی میزان نقدشوندگی بازار بورس ایران و انتخاب معیار مناسب با توجه به بازار غیرنقد بورس اوراق بهادار تهران

ج . مطالعه در زمینه داده‌کاوی، انواع تکنیک‌های موجود در این حوزه و بررسی الگوریتم و نحوه‌ی عملکرد تئوری مجموعه راف

د . مشخص نمودن متغیرهای شرطی و تعیین معیارهای کلان اقتصادی

ه . جمع‌آوری داده‌ها از پایگاه‌های داده مختلف، یکپارچه‌سازی و آماده‌سازی داده‌های گردآوری شده، طراحی جدول‌های داده با استفاده از نرم‌افزار EXCEL

ی . اجرای فرآیند تکنیک راف ست با استفاده از نرم‌افزار ROSETTA و دریافت خروجی‌ها و قوانین مربوطه

ط. ارزیابی میزان دقت مدل به دست آمده با استفاده از تکنیک Cross Validation

ظ. تجزیه و تحلیل نتایج حاصل از تئوری راف ست و بررسی ارتباط میان متغیرهای شرطی و متغیر تصمیم

۱-۸ قلمرو تحقیق :

قلمرو موضوعی

تحقیق حاضر از منظر چارچوب موضوعی شامل بازده سهام، نقدشوندگی سهام، متغیرهای کلان اقتصادی، تکنیک‌های داده‌کاوی و بررسی الگوهای رفتاری پنهان می‌باشد و با توجه به گرایش روزافزون سرمایه‌گذاران به بازارهای سرمایه، نتایج حاصل از این پژوهش می‌تواند در اخذ تصمیمات سرمایه‌گذاری و هم‌چنین تخصیص بهینه منابع مالی مفید و مؤثر می‌باشد.

قلمرو مکانی

قلمرو مکانی تحقیق حاضر سهام شرکت سایپا، یکی از شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌باشد.

قلمرو زمانی

محدوده‌ی زمانی مورد بررسی در این پژوهش اطلاعات بورس اوراق بهادار تهران طی سال‌های ۱۳۸۸ تا ۱۳۹۲ می‌باشد.

۱-۹ نوآوری تحقیق :

مبحث پیش‌بینی بازده سهام، به دلیل اهمیت‌های بی‌شماری که دارد، سال‌هاست که مورد توجه محققان مالی، سرمایه‌گذاران و تحلیل‌گران است و تلاش‌های بسیاری در زمینه ارائه مدلی جامع و فراگیر انجام شده است. از مدل قیمت‌گذاری دارایی‌های مالی آغاز شده و به مرور تکامل یافته است و با توجه به ویژگی‌های ذاتی آن نظیر آشوب‌گونه بودن و ... ، استفاده از مدل‌های سری زمانی غیرخطی نسبت به مدل‌های خطی در مدل‌سازی متغیرهای مالی، در دهه گذشته افزایش چشم‌گیری داشته است. بسیاری از مطالعات در این زمینه نشان داده‌اند، مدل‌های غیرخطی قدرت تخمین بالاتری را نسبت به مدل‌های خطی دارند و قادرند رفتار بازده دارایی‌ها را با درصد اطمینان بالاتری مدل‌سازی نمایند. در میان

مدل‌های غیرخطی، شبکه‌های عصبی مصنوعی از استقبال زیادی در این حیطه برخوردار بودند و مدل‌های بی‌شماری با معماری‌های متفاوت ارائه شده است. اما این روش عیوب و نواقص بسیاری دارد که مهم‌ترین آن‌ها عبارت‌اند از :

قواعد یا دستورات مشخصی برای طراحی شبکه جهت یک کاربرد اختیاری وجود ندارد.

در مورد مسائل مدل‌سازی، صرفاً نمی‌توان با استفاده از شبکه عصبی به فیزیک مساله پی‌برد؛ به عبارت دیگر مرتبط ساختن پارامترها یا ساختار شبکه به پارامترهای فرآیند معمولاً غیر ممکن است.

دقت نتایج بستگی زیادی به اندازه مجموعه آموزش دارد.

پیش‌بینی عملکرد آینده شبکه (عمومیت یافتن آن) به سادگی امکان پذیر نیست.

علاوه بر معایب کلی ذکر شده در بالا، این روش محدودیت‌هایی در زمینه پیش‌بینی بازده سهام دارد که بدین شرح‌اند:

هر چند سری زمانی بازده روزانه شرکت‌های مورد بررسی یک فرآیند غیرتصادفی است ولی پیچیدگی‌های بسیار زیادی دارد و هنگامی که از شبکه‌های عصبی جهت پیش‌بینی بازده استفاده می‌شود، طراحی مدل شبکه عصبی نیازمند بکارگیری شبکه با تعداد لایه‌ها و نرون‌های میانی نسبتاً زیادی می‌باشد.

هرچند معمولاً افزایش تعداد لایه‌ها و نرون‌های میانی منجر به افزایش میزان دقت شبکه در تخمین می‌گردد، ولی این میزان افزایش تا حدی معقول است، به طوری که بعد از آن، منجر به کاهش میزان کارائی شبکه می‌گردد.

ساختار خاصی برای شبکه عصبی جهت پیش‌بینی بازده کلیه شرکت‌ها نمی‌توان طراحی نمود، زیرا هر سری زمانی با ساختار متفاوتی به جواب بهینه نزدیک می‌گردد[۱۲].

نوآوری این پژوهش استفاده از تکنیک پیشرفته‌ی راف ست، یکی از روش‌های محاسبات نرم است که به صورت تلفیقی با متدهای دیگر، از جمله سیستم فازی و یا شبکه‌های عصبی مصنوعی، بکار گرفته می‌شود. روش‌های تلفیقی به نسبت روش‌های ساده از قدرت بیشتری برخوردارند، چرا که روند اصلی خود را با نقاط قوت سایر روش‌ها تقویت می‌کنند و نسلی تکامل یافته‌تر با دقت بالاتر به وجود می‌آورند. تئوری راف ست تلفیقی، توانایی دخالت دادن داده‌های کیفی، حذف سریع داده‌های حشو، طبقه‌بندی و مرتب‌سازی داده‌ها و استخراج قواعد و الگوهای نظام‌مند از دل پایگاه داده را دارا می‌باشد. در این پژوهش تئوری راف ست با سیستم فازی تلفیق شده است.

۱-۱۰ تعریف مختصر اصطلاحات، مفاهیم و واژه‌های کلیدی

بازده سهام: بازده عبارت است از نسبت کل عایدی (ضرر) حاصل از ریسک سرمایه‌گذاری در یک دوره معین به مقدار سرمایه‌ای که برای کسب این عایدی در اول همان دوره بکار گرفته و مصروف گردیده است. عایدی یک سرمایه‌گذاری از مجموع دو عامل سود دریافتی[۴] و سود یا زیان سرمایه[۵] حاصل می‌گردد.

نقدشوندگی سهام: قابلیت نقدشوندگی که میزان نزدیکی دارایی مالی به پول نقد را بیان می‌کند، به صورت امکان انجام معاملات به سرعت، با هزینه اندک و بدون تحت تاثیر قرار دادن شدید قیمت تعریف می‌شود.

داده‌کاوی: داده‌کاوی عبارت است از فرآیند یافتن دانش از مقادیر عظیم داده‌های ذخیره شده در پایگاه داده انبار داده و یا دیگر مخازن اطلاعات. در یک تعریف غیررسمی داده‌کاوی فرآیندی است خودکار برای

[۴] : Yield
[۵] : Capital Gain (Loss)

استخراج الگوهایی که دانش را بازنمایی می‌کنند، که این دانش به صورت ضمنی در پایگاه داده‌های عظیم، انبار داده و دیگر مخازن بزرگ اطلاعات، ذخیره شده است.

تئوری مجموعه راف: یکی از تکنیک‌های داده‌کاوی است و به منظور استخراج قوانین و مقررات حاکم بر داده‌های انبوه، پیش‌بینی و طبقه‌بندی آن‌ها و ارزیابی اهمیت داده‌ها مورد استفاده قرار می‌گیرد. هدف اصلی از تحلیل مجموعه راف به دست آوردن مفاهیم تقریبی از داده‌های اکتسابی می‌باشد. این تئوری یک ابزار قدرتمند ریاضی برای استدلال در موارد ابهام و نایقینی است که روش‌هایی را برای زدودن و کاستن اطلاعات و دانش نامربوط یا مازاد بر نیاز از پایگاه‌های داده مهیا می‌سازد.

فصل دوم

مبانی نظری و ادبیات تحقیق

۲-۱ بخش اول : مفاهیم بنیادی

۲-۱-۱ مقدمه :

امروزه با توجه به اهمیت و گسترش روز افزون بازارهای سرمایه در تجهیز و جمع‌آوری سرمایه‌های کوچک فردی به سمت فعالیت‌های تولیدی، شناسایی رفتار سرمایه‌گذاران و متغیرهای تأثیرگذار بر قیمت و بازده سهام در این بازارها اهمیت زیادی پیدا کرده است. بدیهی است که سرمایه‌گذاری در بورس، بخش مهمی از اقتصاد کشور را تشکیل می‌دهد و بدون شک بیشترین میزان سرمایه از طریق بازارهای بورس در سرتاسر جهان مبادله می‌شود و اقتصاد ملی به شدت متأثر از عملکرد این بازار می‌باشد. علاوه بر آن، بازار بورس هم برای سرمایه‌گذاران حرفه‌ایی و هم برای عموم مردم یک ابزار سرمایه‌گذاری در دسترس محسوب می‌شود. بازارهای بورس از پارامترهای کلان اقتصادی و غیراقتصادی و متغیرهای بسیار دیگری تأثیر می‌پذیرند، متعدد بودن عوامل مؤثر بر بازارهای سرمایه و ناشناخته بودن آن‌ها، موجب عدم اطمینان در زمینه سرمایه‌گذاری شده است، بنابراین محققان سعی دارند تا با استفاده از تکنیک‌ها و روش‌های مختلف، فاکتورهای مهم و مؤثر در سرمایه‌گذاری را شناسایی و پیش‌بینی کرده و تا آن‌جا که امکان دارد درصد اطمینان از بازده سرمایه‌گذاری را افزایش دهند.

۲-۱-۲ بازارهای مالی و طبقه‌بندی آن‌ها :

در هر اقتصاد چهار بازار اصلی وجود دارد که عبارتند از : بازار کالا، بازار کار، بازار پول و بازار سرمایه. بازارهای کالا و کار، مجموعاً بخش واقعی اقتصاد و بازارهای پول و سرمایه، بخش مالی اقتصاد (بازارهای مالی) نامیده می‌شوند. این دو بخش به مثابه چرخ‌های دوچرخه‌ایی هستند که باید همگام با هم حرکت کنند تا منجر به رشد و توسعه اقتصادی گردند. بازار مالی بازاری است که دارایی‌های مالی در آن مبادله (خرید و فروش) می‌شوند. هر چند وجود بازار مالی شرط ضروری برای ایجاد و مبادله دارایی مالی نمی‌باشد، در بیشتر اقتصادها ایجاد و مبادله دارایی‌های مالی در نوعی بازار مالی صورت می‌پذیرد.

طبقه‌بندی بازارهای مالی:

طبقه‌بندی بازارهای مالی از جنبه‌های مختلفی انجام می‌گیرد. یکی از این طبقه‌بندی‌ها بر حسب سررسید حق مالی است، برای مثال بازاری برای ابزار بدهی کوتاه مدت وجود دارد که آن را بازار پول می‌نامند و بازاری که در آن دارایی مالی با سررسید طولانی‌تر معامله می‌شود، بازار سرمایه نام دارد. بازارهای مالی را از لحاظ عمر ابزار مورد معامله نیز می‌توان دسته‌بندی نمود که در این حالت دو نوع بازار اولیه و بازار ثانویه وجود دارد. روش دیگر طبقه‌بندی بازارها بر حسب حق مالی است که بازارهای بدهی (اوراق قرضه) و بازارهای سهام را از یکدیگر مجزا می‌کند.

بازار پول : بازار پول، بازار مطالبات کوتاه مدت (حداکثر تا یک سال) را در برمی‌گیرد. گرایش اصلی بازار پول اصولاً بر ابزارهایی است که با استفاده از آن‌ها بتوان به سرعت وضعیت نقدشوندگی را تغییر داد.

بازار سرمایه : بازار سرمایه، بازاری است که ارتباط تنگاتنگی با پس‌انداز و سرمایه‌گذاری دارد و به مثابه یک واسطه‌ی مالی، پس‌انداز واحدهای اقتصادی دارای مازاد را به واحدهای دارای کسری، برای سرمایه‌گذاری انتقال می‌دهد. به عبارت دیگر بازار سرمایه، نقش هدایت و تخصیص منابع اقتصادی را به عهده دارد. در این بازار نیازهای بلند مدت (معمولاً بیش از یک سال) تأمین می‌شود. یکی از ابزارهای مالی رایج در بازار سرمایه، سهام می‌باشد.

بازار اولیه : وجه مشخصه بازار اولیه این است که در آن، شرکت‌کنندگان در بازار (پس‌اندازکنندگان و سرمایه‌گذاران) دارایی‌های مالی را مستقیماً از منابع اولیه آن خریداری می‌کنند. بازارهای اولیه، واحدهای اقتصادی را قادر می‌سازند که نیازهای مالی خود را از طریق مراجعه به عموم و انتشار سهام تأمین نمایند.

بازار ثانویه : بازار ثانویه بازاری است که در آن، معامله بر روی اوراق بهادار موجود، که قبلاً انتشار یافته‌اند صورت می‌گیرد. در این بازار قیمت‌ها بر اساس مکانیزم عرضه و تقاضا تعیین می‌شوند و شرایط لازم جهت تبدیل دارایی‌های مالی به وجه نقد فراهم می‌شود.

در بازار سهام که از زیر مجموعه‌های بازار ثانویه محسوب می‌شود، سهام عادی، سهام ممتاز و اوراق مورد معامله قرار می‌گیرند. در بعضی از بازارهای ثانویه، سهام به صورت بازارهای حراج[1] است. این بازارها شامل فرآیند حراج (مزایده) در یک مکان فیزیکی معین است. یکی دیگر از اشکال بازارهای ثانویه سهام، بازار مبتنی بر چانه‌زنی[2] است که شامل شبکه‌ایی از معامله‌گران است که از طریق آمادگی برای خرید و فروش اوراق بهادار با قیمت‌های مشخص، بازار را تشکیل می‌دهند. در این نوع از بازارها اوراق بهادار پذیرفته نشده در بورس مورد معامله قرار می‌گیرند.

۲-۱-۳ تعریف سهام:

سهم قسمتی از سرمایه شرکت سهامی است که مشخص‌کننده میزان مشارکت و تعهدات و منافع صاحب آن در شرکت سهامی می‌باشد. ورقه سهم سند قابل معامله‌ای است که نماینده تعداد سهامی است که صاحب آن در شرکت سهامی دارد. سهم ممکن است با نام و یا بی‌نام باشد. در صورتی که برای برخی از سهام شرکت با رعایت مقررات قانون تجارت مزایایی قائل شوند این گونه سهام، سهام ممتاز نامیده می‌شود.

مجموع ارزش اسمی سهام، سرمایه شرکت را تشکیل می‌دهد و پس از انحلال شرکت هر سهامداری حق دارد مبلغ اسمی سهام خود را دریافت نماید. در صورتی که سهم انتفاعی جز سرمایه شرکت نباشد، صاحب آن حق مطالبه مبلغ اسمی سهام را پس از انحلال شرکت ندارد.

اصولاً تمامی سهامداران در شرکت سهامی از تساوی حقوق برخوردارند و این حقوق شامل میزان مشارکت تعهدات و منافع صاحب سهم در شرکت سهامی می‌باشد و برای استفاده از این حقوق دارنده سهم باید مبلغ اسمی سهمی را که پذیره‌نویسی نموده در مهلت قانونی آن پرداخت نماید.

---

[11] : Auction Market
[2] : Negotiated Market

انواع سهام

سهام با نام : سهمی است که بر مالکیت شخص معینی دلالت می‌کند و مشخصات صاحب سهم در ورقه سهم و در دفتر ثبت شرکت درج می‌گردد. در موارد زیر صدور سهام با نام الزامی می‌باشد:

۱- مادامی که تمام مبلغ اسمی هر سهم بی‌نام پرداخت نشده باشد.

۲- سهامی که مدیران شرکت به عنوان وثیقه می‌سپارند و تا خاتمه حساب دوره تصدی مدیران غیرقابل انتقال می‌باشد.

۳- سهام محجورین و صغار به منظور حمایت از حقوق آنها.

سهام با نام و هر گونه نقل و انتقال آن باید در دفتر ثبت سهام شرکت ثبت گردد، در غیر این صورت فاقد اعتبار قانونی برای شرکت و اشخاص ثالث خواهد بود.

سهام بی نام : سهمی است که نام دارنده در آن درج نمی‌شود و به صورت سند در وجه حامل، دارنده آن مالک شناخته می‌شود تا زمانی که خلاف آن ثابت گردد. این سهم به راحتی قابل نقل و انتقال بوده و انتقال مالکیت با قبض و اقباض تحقق می‌یابد. این نوع سهام از حیث صلاحیت محاکم در حکم اموال منقول است و باید به دادگاه محلی رجوع شود. سرقت و یا مفقود شدن سهام بی نام و هم‌چنین خیانت در امانت امین و وکیل نسبت به سهام بی نام که به آنها سپرده شده، خطر بزرگی برای صاحب آن محسوب می‌شود. زیرا بی نام بودن سهم مانع شناسایی دارنده اصلی می‌شود و سارق و یا یابنده به آسانی می‌تواند سهم مزبور را در بازار بورس و یا خارج از آن به اشخاص دیگر بفروشد و خریدار نیز حق دارد به تصرف خود به عنوان مالکیت استناد نماید.

سهام ممتاز : یکی از اصول اساسی شرکت‌های سهامی تساوی حقوق سهامداران است؛ این اصل در دو مورد نقض و منجر به صدور سهام ممتاز می‌شود، اول آن‌که شرکت با مشکلات مالی مواجه باشد و نتواند به آسانی به منابع مالی دسترسی یابد. در این صورت ناچار خواهد بود امتیازات بیشتری برای سهامداران

جدید قائل شود و دوم آن‌که در هنگام افزایش سرمایه شرکت، شرکت از ذخایر مالی خود برای نیل به این هدف استفاده نماید که در این صورت باید حقوق بیشتری برای سهامداران سابق شرکت در نظر گرفته شود.

امتیازاتی که توسط سهام ممتاز به وجود می‌آید می‌تواند به صورت برداشت نفع ثابت از سود قابل تقسیم شرکت باشد و یا این‌که دارنده سهم ممتاز می‌تواند نسبت به صاحبان سهام عادی در برداشت ارزش اسمی سهام ممتاز به هنگام انحلال و تصفیه و تقسیم دارائی شرکت اولویت داشته باشد. به هنگام انحلال و تصفیه شرکت، مطالبات معوقه صاحبان سهام ممتاز از دارائی شرکت پرداخت می‌گردد و نیز هنگام استهلاک سهام شرکت، صاحبان سهام ممتاز دارای اولویت می‌باشند.

سهام مؤسس یا انتفاعی : سهم مؤسس یا انتفاعی سهمی است که دارنده آن بدون داشتن حقی به سرمایه شرکت از منافع شرکت به نسبت سهم خود بهره‌مند می‌گردد و معمولاً این گونه سهام را مؤسسان بابت اطلاعات علمی، فنی و تجاری‌ای که دارند، به خود تخصیص می‌دهند. در اساسنامه شرکت می‌بایستی حقوق صاحبان سهام مؤسس یا انتفاعی مثل حق حضور در مجامع حق رأی، حق اولویت در پذیره نویسی سهام و غیره تصریح گردد، زیرا سهامداران مزبور مالک سرمایه نمی‌باشند و نمی‌توانند سهامداران شرکت محسوب شوند ولی در منافع شرکت حقوقی دارند که از آن مانند طلبکار منتفع می‌گردند و استفاده آن‌ها از سود شرکت بستگی به موفقیت و سودآوری شرکت دارد.

۲-۱-۴ بازده سهام :

به منفعت و سودی که از یک سرمایه‌گذاری حاصل می‌شود بازده می‌گویند. سرمایه‌گذاری عبارت است از تخصیص منابع به دارایی‌های واقعی نظیر زمین و خانه و دارایی‌های مالی نظیر اوراق بهادار که میزان بازده آن متناسب با ریسک مورد انتظار باشد. بازده در فرآیند سرمایه‌گذاری نیروی محرکی است که ایجاد انگیزه می‌کند و پاداشی برای سرمایه‌گذاران محسوب می‌شود. بازده ناشی از سرمایه‌گذاری برای

سرمایه‌گذاران حائز اهمیت است، زیرا کل فرآیند بازی سرمایه‌گذاری به منظور کسب بازده انجام می‌گیرد. ارزیابی از بازده سرمایه‌گذاری تنها راه منطقی (قبل از ارزیابی ریسک) برای سرمایه‌گذاران به منظور مقایسه سرمایه‌گذاری‌های جایگزین و متفاوت می‌باشد.

مبحث بازده به دو بحث بازده تحقق یافته و بازده مورد انتظار تقسیم می‌شود. بازده مورد انتظار[1] عبارت است از بازده تخمینی یک دارایی که سرمایه‌گذاران انتظار دارند در یک دوره آینده بدست آورند. این بازده که با عدم اطمینان همراه است، امکان دارد تحقق یابد و یا اصلاً محقق نگردد. سرمایه‌گذاران برای کسب بازده مورد انتظار بایستی یک نوع دارایی را خریداری کنند و توجه داشته باشند که این بازده ممکن است تحقق نیابد. سرمایه‌گذاری بر روی اوراق بهادار ریسک‌دار و بلند مدت می‌تواند باعث برآورده شدن بازده مورد انتظار سرمایه‌گذاران شود، در حالی که در کوتاه‌مدت این امر کمتر اتفاق می‌افتد.

بازده عبارت است از نسبت کل عایدی (ضرر) حاصل از ریسک سرمایه‌گذاری در یک دوره معین به مقدار سرمایه‌ای که برای کسب این عایدی در اول همان دوره بکار گرفته و مصروف گردیده است. عایدی یک سرمایه‌گذاری از دو طریق ۱) سود دریافتی[2]: مهم‌ترین جز بازده، سودی است که به صورت جریانات نقدی دوره‌ای سرمایه‌گذاری بوده و می‌تواند به شکل بهره یا سود تقسیمی باشد. ویژگی ممتاز این دریافت‌ها این است که منتشر کننده پرداخت‌هایی را به صورت نقدی به دارنده دارایی پرداخت می‌کند. این جریانات نقدی با قیمت اوراق بهادار نیز مرتبط است. ۲) سود یا زیان سرمایه[3]: دومین جز مهم بازده، سود (زیان) سرمایه است که مخصوص سهام عادی است ولی در مورد اوراق قرضه بلند مدت و سایر اوراق بهادار با درآمد ثابت نیز مصداق دارد. به این جز که ناشی از افزایش (کاهش) قیمت دارایی است سود (زیان) سرمایه می‌گویند. این سود (زیان) سرمایه ناشی از اختلاف بین قیمت خرید و قیمت در زمانی است که دارنده اوراق قصد فروش آن‌ها را دارد. این اختلاف می‌تواند سود و یا زیان باشد.

---

[1] : Expected Return
[2] : Yield
[3] : Capital Gain (Loss)

بازده سرمایه‌گذاری شامل بازده کمی و کیفی می‌باشد. بازده کمی همان سود حسابداری، مثل EPS و منفعت سرمایه است و بازده کیفی اعتبار شرکت، مثل سرقفلی و امتیازات اجتماعی است. در تئوری بازده سرمایه‌گذاری‌ها (قسمت کمی) با استفاده از رابطه (۲-۱) محاسبه می‌گردد:

رابطه ۲-۱

$$r_e = \frac{D_1}{P_0} + \frac{xP_1 - P_0}{P_0}$$

در رابطه (۲-۱)، r_e بازده مورد انتظار، D_1 سود حسابداری است که عاید سرمایه‌گذار خواهد شد، P_0 ارزش سرمایه در زمان $t=0$، P_1 ارزش سرمایه در زمان $t=1$ و x ضریب تبدیل سهام می‌باشد.

۲-۱-۵ نقدشوندگی سهام:

یکی از کارکردهای مهم بازارهای مالی و به ویژه بورس اوراق بهادار، تسهیل، تسریع و کاهش هزینه در فرآیند تبدیل دارایی مالی به وجه نقد و برعکس یعنی تبدیل وجه نقد به دارایی مالی است، که از این ویژگی به نقدشوندگی یاد می‌شود. نقدشوندگی قابلیت خرید و فروش مقادیر قابل توجهی از اوراق بهادار به سرعت و با تأثیر بسیار اندک در قیمت است. این ویژگی که هم در سطح اوراق بهادار انفرادی مانند یک سهم و هم در سطح کل بازار سهام قابل بررسی است، نقش مهمی در ارزش‌گذاری دارایی‌های مالی ایفا می‌کند به‌طوری که در شرایط مساوی، اوراق بهادار با نقدشوندگی بالاتر، قیمت بیشتری خواهند داشت. سهام دارای قابلیت نقدشوندگی بالا، طبیعتاً ریسک نگهداری کمتری داشته و از جذابیت بیشتری برخوردار می‌باشند در حالی که با کاهش قابلیت نقدشوندگی یک سهم، از جذابیت آن کاسته می‌شود، مگر در شرایطی که بازده بیشتری عاید دارنده آن شود. مسأله نقدشوندگی سهام در بورس اوراق بهادار تهران که به دلیل فقدان ساز و کارهای فراهم آورنده نقدشوندگی، در ردیف بورس‌های غیرنقد جهان قرار می‌گیرد، یکی از دغدغه‌های اصلی سرمایه‌گذاران است زیرا به همان اندازه که بازده بالقوه بالای سهام برای آنان جذاب است، از ریسک نقدشوندگی هراس دارند. ارزیابی میزان نقدشوندگی یک سهم برای سرمایه‌گذاران، علی‌رغم تعریف و مفهوم ساده‌ایی که دارد، کاری بس دشوار است زیرا این ویژگی یک

متغیر تک بعدی نبوده و جنبه‌های مختلف آن، معیارهای متفاوتی دارد.

نقدشوندگی چهار بعد متفاوت (۱) زمان انجام معامله، توانایی انجام معامله به صورت آنی در قیمت جاری بازار، (۲) استحکام، توانایی خرید و فروش دارایی در تقریبا یک قیمت و یک زمان، (۳) عمق، توانایی خرید یا فروش میزان مشخصی از یک دارایی بدون تحت تأثیر قرار دادن قیمت و (۴) انعطاف‌پذیری، سرعت بازگشت قیمت‌ها به سطوح قبلی پس از تغییر در قیمت به واسطه انجام معاملات بزرگ، دارد [۱۳] . ویس[1] (۲۰۰۴)، پس از بررسی روابط همبستگی میان ۳۱ معیار نقدشوندگی در ۱۸ بازار دنیا، ۲۳ معیار مناسب را معرفی کرد. وی دو دسته کلی معیارهای تک بعدی و چندبعدی را ارائه نمود که معیارهای چندبعدی به صورت تلفیقی از چند معیار تک‌بعدی می‌باشند. از میان چهار دسته معیارهای تک بعدی ۱) اندازه شرکت، ۲) حجم معامله، ۳) شکاف قیمت پیشنهادی خرید و فروش و ۴) زمان بین دو معامله انجام شده متوالی، سه دسته آخر به ترتیب پنج، نه و دو معیار را به خود اختصاص می‌دهند، در حالی که تعداد و تنوع معیارهای چندبعدی بسیار زیاد است و هر یک با توجه به ویژگی‌های خاص خود، کاربرد متفاوتی دارد.

پژوهشگران از معیارهای تک‌بعدی بسیاری استفاده کرده‌اند که تمرکز اکثر آن‌ها بر روی ۱)نرخ گردش سهام[2]، ۲) تعداد معاملات و ۳) اختلاف قیمت پیشنهادی خرید و فروش بوده است. این معیارها به ترتیب نماینده دسته‌های حجم معامله، زمان و شکاف قیمت پیشنهادی خرید و فروش هستند. در این میان عامل اندازه شرکت، که دسته نخست را تشکیل داده است، به دلیل فقدان نوسان‌پذیری کافی و ضعف در ارائه نتایج مناسب مورد استفاده قرار نمی‌گیرد [۱۳].

۱) نرخ گردش سهام که به بعد عمق و زمان نقدشوندگی اشاره دارد با استفاده از رابطه (۲-۲) محاسبه می‌گردد:

[1]: Wyss, R.
[2] : Turnover

$$V_t = \sum_{i=1}^{N_t} p_i q_i$$ رابطه (۲-۲)

در این رابطه N_t تعداد معاملات در بازه زمانی $[1,t]$، p_i قیمت هر سهم در معامله i-ام و q_i تعداد سهام معامله شده در معامله i-ام می‌باشد. بالاتر بودن نرخ گردش سهام بیانگر نقدشوندگی بیشتر آن است. حالت تجمعی[1] این نرخ در رابطه (۲-۳) نشان داده شده است. در حالت تجمعی صورت کسر تعداد کل سهام معامله شده و مخرج کسر تعداد کل سهام منتشره می‌باشد و کل عبارت بیان‌گر جریان آزاد هر سهم است که حجم معامله روزانه از سهام منتشره را نشان می‌دهد.

$$AT_{it} = V_{it} / I_i$$ رابطه (۲-۳)

۲) تعداد معاملات در واحد زمان با N_t بیان می‌شود که مقادیر بیشتر آن، نقدشوندگی بهتر سهام را بیان می‌کند.

۳) اختلاف قیمت پیشنهادی خرید و فروش که به آن اسپرد[2] گفته می‌شود با رابطه (۲-۴) محاسبه می‌گردد و بیانگر هزینه معاملاتی است، زیرا P_t^A کمترین قیمت پیشنهادی فروش و P_t^B بیشترین قیمت پیشنهادی خرید را به نمایش می‌گذارند. با کمتر شدن این فاصله و کاهش هزینه معاملاتی، خریدار و فروشنده طی زمان کوتاه‌تری به قیمت تعادلی رسیده و حجم معاملات سهام فزونی می‌یابد که منجر به افزایش نقدشوندگی می‌شود.

$$Sabs_t = P_t^A - P_t^B$$ رابطه (۲-۴)

انتخاب معیارهای چندبعدی کمی دشوارتر است زیرا سازگاری آن‌ها با شرایط حاکم بر بازار، از جمله غیرنقد بودن آن، امری ضروری است و برخی از معیارها تنها برای برخی بازارها مناسب هستند. اکثر معیارهای اندازه‌گیری نقدشوندگی نیاز به تراکنش‌های بالا و داده‌های مربوط به قیمت‌گذاری دارند که در برخی از بازارها، خصوصاً بازارهای نوپا، در دسترس نمی‌باشد. به منظور برطرف کردن این مشکل،

[1] : Aggregate Turnover
[2] : Spread

تحقیقات بسیاری انجام گرفته و معیارهای متفاوتی ارائه شده است.

ژانگ[1](۲۰۱۰) که در پژوهش خود به بررسی معیارهای اندازه‌گیری نقدشوندگی در بازارهای نوپا پرداخته، معتقد است معیارهای آمیوست[2]و گاما[3] برخلاف معیارهای آمیهیود[4]و نسبت حجم صفر[5]، توسط سایر معیارها مغلوب می‌شوند و مناسب نیستند. کیم و لی[6] و هم‌چنین فونگ و همکاران[7] استفاده از دو معیار عدم نقدشوندگی آمیهیود و نسبت حجم صفر را در بازارهای نوپا بلامانع و مطلوب یافتند [۱۴,۱۵,۱۶]، بنابراین افزودن دو معیار پرکاربرد چندبعدی آمیهیود و حجم صفر ،که با ساختار بازار سرمایه ایران سازگار هستند، می‌تواند به نتایج حاصل از بررسی‌ها اطمینان بیشتری بخشد.

معیار آمیهیود که در سال ۲۰۰۲ توسط وی ارائه شده است، نسبت بازده به حجم را از طریق ساختار ساده‌ایی به صورت روزانه به منظور بررسی اثر قیمت به خدمت می‌گیرد. این معیار که از دیدگاه گوینکو و همکاران[8] در زمینه بررسی اثر قیمت، بر معیارهای پاستور[9]، گامای استمبوف[10] و آمیوست ارجح‌تر است، است، ارتباط مثبت قوی‌ای با بازده مورد انتظار سهام دارد و در روزهایی که تغییرات قیمت صفر است نیز قابل محاسبه می‌باشد. این دو مزیت عمده خصوصاً برای بازارهایی که فاقد زیرساخت‌های کلان بازار سرمایه بوده و بازار توسعه یافته‌ایی ندارند جایگاه ویژه‌ایی به معیار آمیهیود بخشیده و کاربرد آن را افزایش داده است.[۱۹,۱۸,۱۷]. این معیار که مقدار کمتر آن بیان‌گر نقدشوندگی بیشتر سهم است، با استفاده از رابطه (۲-۵) اندازه‌گیری می‌شود:

[1] : Huiping Zhang
[2] : Amivest
[3] : Gamma
[4] : Amihud
[5] : Zero vol
[6] : Sunho Kim & Kuan-Hui Lee
[7] : Fong, Kingsley; Holden, Caig W & Trzcinka, Charles A.
[8] : Ruslan Y. Goyenko, Craig W. Holden & Charles A. Trzcinka
[9] : Pastor
[10] : Stambaugh's Gamma

$$ILLIQ_{it} = \frac{1}{D_{it}} \sum_{i=1}^{D_{it}} |R_{itd}| / VOLD_{ivtd}$$ رابطه (۲-۵)

D_{it} تعداد روزهایی از ماه t است که برای سهام i داده موجود است. R_{itd} نشان دهنده بازده سهام i-ام در روز d از ماه t می‌باشد و $VOLD_{itd}$ حجم روزانه مربوطه به پول رایج آن کشور است.

متاسفانه معیار آمیهیود برای روزهای معاملاتی با حجم صفر غیرقابل تعریف می‌باشد، بنابراین نیاز به حضور معیاری دیگر برای پوشش این ضعف احساس می‌شود. لزموند، اوگدن و ترسینکا[1](۱۹۹۹) معیار مبتنی بر بازده صفر را پیشنهاد دادند که نسبت تعداد روزهای بازده صفر سهام در ماه، به تعداد کل روزهای معاملاتی آن سهام در همان ماه است. به اعتقاد لزموند (۲۰۰۵) از آنجا که به دلیل هزینه معاملاتی بالا، معامله‌گر تصمیم به عدم انجام معامله می‌گیرد این معیار نشان‌دهنده ارتباط میان نقدشوندگی و ریسک سیاسی در بازارهای نوپا می‌باشد. معیار بازده صفر برای یک روز با حجم معاملاتی مثبت که به دلیل وجود هزینه‌های معاملاتی بالا بازده صفر کسب کرده است، مناسب نمی‌باشد. بکارت، هاروی و لوندبلاد[2] (۲۰۰۷) به منظور رفع این نقص نسبت مبتنی بر صفر دیگری تحت عنوان حجم صفر ارائه کردند که تعداد روزهای حجم صفر یک سهم در ماه به تعداد کل روزهای معاملاتی آن سهم در ماه است و به ZERO VOL معروف می‌باشد. این نسبت بیان می‌کند که تعدد روزهای بدون معامله برای یک سهم، می‌تواند حاکی از نقدشوندگی پایین آن باشد، بنابراین سهامی که نقدشوندگی کمتری دارند معمولا تعداد روزهای معاملاتی با حجم صفر بیشتری به خود اختصاص می‌دهند [۲۰,۱۵].

ژانگ که یکی از محققان برجسته در زمینه نقدشوندگی می‌باشد، در سال ۲۰۱۴ معیار جدیدی تحت عنوان معیار اصلاح‌شده‌ی عدم‌نقدشوندگی آمیهیود[3] که با نماد AdjILLIQ نشان داده می‌شود را ارائه داد. داد. وی که در پژوهش خود در سال ۲۰۱۰ بدین نتیجه دست یافته بود که معیارهای آمیهیود و نسبت حجم صفر به منظور اندازه‌گیری نقدشوندگی در بازارهای نوظهور، بهترین عملکرد را دارند، دست به

[1] : Lesmond, Ogden and Trzcinka
[2] : Bekaert, Harvey and Lundblad
[3] : Modified Version of the Amihud Illiquidity Measure

تلفیق این تلفیق این دو معیار زد. معیار جدید که برای انواع بازارهای نوظهور به خصوص بازارهایی با تعداد معاملات کم و سهام با نرخ گردش پایین به خوبی عمل کرده است، تلفیق معیار اصلی آمیهیود و فراوانی حجم صفر می‌باشد که همبستگی بسیاری با اختلاف قیمت پیشنهادی خرید و فروش و تأثیر قیمت دارد و با استفاده از رابطه (۲_۶) محاسبه می‌گردد.

$$AdjILLIQ_{i,m} = \left[\ln\left(\frac{1}{N_{i,m}} \sum_{t=1}^{N_{t,m}} \frac{|R_{i,t}|}{Vol_{i,t}} \right) \right] \times (1 + \text{ZeroVol}_{i,m}) \qquad \text{رابطه (۲-۶)}$$

در رابطه (۲-۶)، $N_{i,m}$ تعداد روزهایی از ماه m است که سهام i مورد معامله قرار گرفته و حجم غیرصفر دارد. $|R_{i,t}|$ مقدار بازده سهام i در روز t، $Vol_{i,t}$ حجم ریالی سهام i در روز t و $ZeroVol_{i,m}$ درصد روزهای حجم صفر سهام در ماه m می‌باشد. استفاده از این معیار در بازارهای نوپایی که سیستم معاملاتی آن‌ها اطلاعات مربوط به اسپرد را به ثبت نمی‌رساند، پیشنهاد می‌گردد.

۲-۱-۶ بازار کارا و امکان پیش‌بینی قیمت‌ها:

در اغلب مواقع از یوجین فرانسیس فاما، اقتصاددان آمریکایی و یکی از برندگان نوبل اقتصاد سال ۲۰۱۳ میلادی، به عنوان پدر فرضیه بازار کارا یاد می‌شود که مطالعه در این زمینه را از دوره دکترایش آغاز کرده بود. او در ماه می سال ۱۹۷۰ میلادی مقاله‌ای علمی درباره نظریه مالی با عنوان "بازارهای سرمایه کارا: گذری بر نظریه و کارهای تجربی" منتشر کرد و برای نخستین بار فرضیه بازار کارا را بیان نمود. موضوع کارایی بازار یکی از مباحث مهم در زمینه سرمایه‌گذاری محسوب می‌شود. اکثر اقتصاددانان در حوزه امور مالی اعتقاد دارند که سرمایه را باید به نحو بهینه تخصیص داد. به عبارت دیگر هدف منطقی سیاست دولت‌ها، ترغیب و تشویق استقرار بازار کارایی تخصیصی است که در آن شرکت‌هایی که دارای محتمل‌ترین فرصت‌های سرمایه‌گذاری هستند، به وجوه مورد نیاز دسترسی داشته باشند. مفهومی که از کارایی در اینجا مدنظر قرار می‌گیرد، اشاره به این مسأله دارد که تا چه میزان بازار در تعیین قیمت اوراق

بهادار موفق عمل کرده است. بازاری از نظر تخصیص سرمایه کارا محسوب می‌شود که هم کارای درونی باشد و هم کارای بیرونی.

در یک بازار کارای بیرونی (که کارایی اطلاعاتی نیز نامیده می‌شود)، اطلاعات به سرعت و به طور وسیع منتشر شده و بدین وسیله امکان تعدیل سریع قیمت هر ورقه بهادار در مقابل اطلاعات جدید فراهم می‌شود به طوری که قیمت آن، ارزش سرمایه‌گذاری را منعکس می‌کند. بازار کارای درونی (که کارایی عملیاتی نیز نامیده می‌شود)، بازاری است که در آن کارگزاران و معامله‌گران به صورت منصفانه با یکدیگر رقابت می‌کنند به طوری که هزینه مبادلات کم و سرعت انجام آن زیاد است. به بیان بهتر کارایی درونی بر این موضوع تأکید دارد که کارگزاران و معامله‌گران با هزینه‌ایی کم و به سرعت معامله را برای مشتری اجرا کنند. اصطلاح کارایی بازار معمولاً به کارایی بیرونی بازار دلالت دارد.

بازار کارا ویژگی‌های زیر را دارد:

۱.تعداد زیادی سرمایه‌گذار منطقی که به دنبال حداکثر کردن سود خود می‌باشند، به طور فعال از طریق تجزیه و تحلیل، ارزشیابی و معامله سهام در بازار مشارکت دارند. این سرمایه‌گذاران به تنهایی نمی‌توانند بر قیمت اوراق بهادار تأثیر بگذارند.

۲. کلیه سرمایه‌گذاران بدون متحمل شدن هزینه‌ایی به اطلاعات جاری موجود در مورد آینده دسترسی دارند.

۳. کلیه سرمایه‌گذاران قیمت‌های بازار را به دقت زیر نظر دارند و به طور مناسبی اوراق بهادار نگهداری شده را تعدیل می‌کنند.

۴. اطلاعات به صورت تصادفی به وجود می‌آید و مستقل از اعلان سایر اطلاعات می‌باشد.

۵. سرمایه‌گذاران در قبال ارائه اطلاعات جدید سریعاً عکس‌العمل نشان می‌دهند که این خود باعث تغییر قیمت سهام می‌شود.

در حقیقت می‌توان بیان کرد که بازار کارا بازاری است که در آن قیمت اوراق بهادار در کلیه زمان‌ها با ارزش سرمایه‌گذاری آن برابر است. مفهوم بازار کارا بر این فرض استوار است که سرمایه‌گذاران در تصمیمات خرید و فروش خود، تمامی اطلاعات مربوط را در قیمت سهام لحاظ خواهند کرد. بنابراین قیمت فعلی سهام شامل تمامی اطلاعات شناخته شده اعم از اطلاعات گذشته (مانند سود مربوط به فصل یا سال گذشته) و اطلاعات فعلی است.

اشکال مختلف کارایی بازار:

شکل ضعیف کارایی :

اگر به دست آوردن سودهای غیرنرمال (به جز آن‌هایی که به طور شانسی به دست می‌آیند) با استفاده از قیمت‌های گذشته برای تصمیم‌گیری در زمان خرید و فروش اوراق بهادار امکان‌پذیر نباشد، عنوان می‌شود که بازار دارای شکل ضعیف کارایی است. به عبارتی می‌توان گفت در این نوع از کارایی بازار تغییرات گذشته قیمت باید رابطه‌ایی با تغییرات آتی قیمت نداشته باشد. به عبارت دیگر، تغییرات قیمت سهام در طول زمان باید مستقل و یا تقریباً این‌گونه باشد. مفهوم شکل ضعیف فرضیه بازار کارا این است که اطلاعات قیمتی تاریخ گذشته در ارزیابی آتی قیمت تأثیر ندارد.

شکل نیمه قوی کارایی:

اگر کسب سودهای غیرنرمال (بجز آن‌هایی که به طور شانسی به دست می‌آیند) از طریق استفاده از اطلاعات در دسترس عموم برای تصمیم‌گیری در مورد خرید و فروش اوراق بهادار امکان‌پذیر نباشد. عنوان می‌شود که بازار دارای شکل نیمه‌قوی کارایی است. این سطح که سطح خیلی متداول کارایی بازار می‌باشد نه تنها شامل اطلاعات موجود قیمتی است، بلکه شامل تمام اطلاعات شناخته شده و در دسترس مانند اطلاعات مربوط به درآمد، سود تقسیمی، اعلان تجزیه سهام، پیشرفت‌های جدید در مورد

محصولات، مشکلات تأمین مالی و تغییرات مربوط به حسابداری نیز می‌شود. بازار کارای نیمه‌قوی در برگیرنده شکل ضعیف بازار کارا نیز هست و شامل سرمایه‌گذارانی است که از اطلاعات منتشره جدید استفاده نمی‌کنند و این سرمایه‌گذاران انتظار دارند با توجه به متوسط ریسک، بازده بیشتری بدست آورند.

شکل قوی کارایی:

اگر کسب سودهای غیرنرمال (به جز آن‌هایی که به طور شانسی به دست می‌آیند) از طریق استفاده از هر گونه اطلاعات برای تصمیم‌گیری در مورد خرید و فروش اوراق بهادار به هیچ‌وجه امکان‌پذیر نباشد، گفته می‌شود که بازار دارای شکل قوی کارایی است. شکل قوی بازار کارار در برگیرنده شکل ضعیف و نیمه‌قوی نیز هست و نشان‌دهنده بالاترین سطح کارایی بازار است.

امکان پیش‌بینی قیمت‌ها در بازار:

تغییرات قیمت اوراق بهادار در بازار کارا به صورتی تصادفی است. در بازار کارا تمام سرمایه‌گذاران فوراً با ورود اطلاعات جدید واکنش نشان می‌دهند و قیمت سهام با این واکنش تغییر خواهد نمود. در این بازار اطلاعات جدید خبری شگفت‌انگیز است (اطلاعاتی که شگفت‌انگیز نیست قابل پیش‌بینی بوده، قبلاً پیش‌بینی شده است و اکنون کهنه تلقی می‌شود). اخبار شگفت‌انگیز می‌توانند بد یا خوب باشند، در بازار کارا اخبار خوب منجر به افزایش قیمت و اخبار بد منجر به سقوط قیمت خواهد شد. با اعلام افزایش در بازدهی سهام (برای مثال، افزایش در سود نقدی سهام)، قیمت سهام آن افزایش می‌یابد. در بازار کارا هرگونه افزایش یا کاهش در میزان بازدهی غیرقابل پیش‌بینی است. بنابراین تغییرات قیمت در بازار کارا کاملاً تصادفی است. تصادفی بودن تغییرات قیمت به معنای غیرمنطقی بودن قیمت فعلی نیست. به دلیل ورود تصادفی اطلاعات، تغییرات قیمت به عنوان برآیند ورود اطلاعات تصادفی (که گاهی مثبت و گاهی منفی) است، تلقی می‌شود. به هر حال تغییرات در قیمت، نتیجه ارزیابی مجدد سرمایه‌گذاران از وضعیت آتی سهام و اخذ تصمیمات جدید در مورد خرید یا فروش سهام است. از این رو، تغییرات در قیمت سهام اگر چه تصادفی است، اما منطقی است.

مدیریت مالی جدید با بررسی تئوری بازار کارا، وجود ناهنجاری‌هایی را در بازار کارا مورد تأکید قرار داده است. دو دسته ناهنجاری‌های غیرتقویمی[۱] و تقویمی در تئوری بازار کارا بیان شده است.

ناهنجاری‌های غیرتقویمی بی‌قاعدگی‌ها و بی‌نظمی‌هایی هستند که فرضیه بازار کارا را زیر سؤال برده و نمی‌توان آن‌ها را در قالب بی‌نظمی‌های فصلی طبقه‌بندی کرد. در این بی‌قاعدگی‌ها عامل زمان برهم زننده‌ی بازار کارا نیست. ناهنجاری‌های غیرتقویمی به پنج دسته اثر تجزیه سهام، اثر عرضه‌ی اولیه اوراق، اثر شاخص، اثر معاملات محرمانه و اثر تأخیر در ارائه‌ی گزارش سودآوری تقسیم می‌شوند.

اثر تجزیه سهام[۲]: به طور منطقی تجزیه سهام یک شرکت نباید اثری بر ارزش شرکت فوق بگذارد، لیکن شواهد و مطالعات فراوانی حاکی از آن است که تجزیه سهام یک بنگاه سبب افزایش قیمت سهام هم قبل و هم بعد از انتشار اعلامیه‌ی تجزیه‌ی سهام می‌شود. این اثر را "فاما، ادسای و جاین[۳]" مورد مطالعه قرار داده‌اند.

اثر عرضه اولیه اوراق[۴] : ظریف فرد و مهرجو در سال ۱۳۸۳ با تحقیقی تحت عنوان "قیمت‌گذاری سهام شرکت‌ها در اولین عرضه سهام آن‌ها به بورس" بازده کوتاه مدت سهام ۹۱ شرکت جدیدالورود در طول سال‌های ۱۳۷۸ تا ۱۳۸۲ را بررسی کردند. نتایج این تحقیق نشان داد بازده کوتاه مدت این شرکت‌ها بیشتر از بازده بازار بوده است.

اثر شاخص[۵] : جنگ هاون[۶] در سال ۲۰۰۵ با بررسی شاخص S&P 500 و نحوه تغییر قیمت سهام شرکت‌هایی که در این شاخص قرار می‌گیرند و یا خارج می‌شوند، اظهار کرد که قیمت سهام شرکت‌هایی که به شاخصی همچون S&P 500 اضافه شوند به سرعت افزایش می‌یابد و بازده غیرمعمول کسب می‌کنند.

---

[1] : Non Calender Anomalies
[2] : Stock Split Effect
[3] : Fama, Desai & Jain
[4] : Initial Public Offering Effect
[5] : Index Effect
[6] : Jong Hwan

اثر معاملات محرمانه[1] : افرادی نظیر مدیران شرکت، سهامداران عمده و برخی از افراد حرفه‌ای بازار، به دلیل عدم کارایی اطلاعاتی بازار، اطلاعات بروزتری نسبت به شرکت دارند. مطالعات زیادی همچون فیزتی (۱۹۷۶)، سیهرن (۱۹۸۶)، لین هار (۱۹۹۰) و لاکونیشوک (۲۰۰۱) نشان دادند که محرمین اطلاعات شرکت اغلب می‌توانند تغییرات قیمت را پیش‌بینی کنند و بازده‌های غیرعادی نسبت به بازده بازار کسب کنند.

اثر تأخیر در ارائه‌ی گزارش سودآوری[2] : برخی محققان معتقدند شرکت‌هایی که گزارش سودهای خود را به سایر شرکت‌ها با تأخیر ارائه می‌دهند، در دوره زمانی قبل از ارائه‌ی گزارش با بازده‌های منفی مواجه می‌شوند. هرچند از نظر سودآوری تغییرات عمده‌ای در شرکت صورت نپذیرفته باشد.

ناهنجاری‌های تقویمی : در مورد الگوها و بی‌قاعدگی‌های فصلی در بازارهای مالی از جمله بازار سهام و اوراق قرضه، شواهد و مدرک زیادی از نیم‌قرن گذشته تاکنون در دست است. مطابق فرضیه بازار کارا، قیمت سهام در یک بازار کارا همواره به صورت تصادفی تغییر می‌یابد که دلیل آن نیز پاسخ مثبت سهام به اطلاعاتی است که به گونه‌ایی تصادفی در طول زمان منتشر و عرضه می‌شود. حال اگر زمان، عامل تغییردهنده قیمت سهام باشند، به گونه‌ای که در طول دوره‌ی خاص، علاوه بر اطلاعات عرضه شده (به صورت تصادفی)، زمان نیز بر تغییر قیمت سهام مؤثر باشد و ماهیت تصادفی بودن رفتار بازار را تغییر می‌دهد. به این گونه اثرات، اثرات تقویمی یا بی‌قاعدگی‌های تقویمی گفته می‌شود. این ناهنجاری‌ها به پنج دسته اثر روزهای آخر هفته، اثر ماه‌های خاص سال در تقویم اسلامی (ماه‌های رمضان، محرم و صفر)، اثر روزهای تعطیل با اثر قبل از روزهای تعطیل، اثر آخر ماه و بین ماه و اثر پایان سال (اثر ژانویه، اثر دسامبر).

اثر روزهای آخر هفته : اثر روزهای هفته به وجود الگوهایی در بازدهی سهام در گذشته اشاره دارد که این الگو به روزهای ویژه هفته مربوط می‌شود. چنین روابطی به طور عمده در آمریکا تصدیق شده است. به

[1] : Insider Transaction Effect
[2] : Late Earning Reporter Effect

نحوی که آخرین روزهای معاملاتی هفته با بازدهی غیرعادی مثبت همراه است. در حالی که دوشنبه یعنی اولین روز معاملاتی هفته با بازدهی کمتر از روزهای دیگر و حتی بازدهی منفی مشخص می‌شود. وجود چنین اثری خود می‌تواند شاهدی بر این ادعا باشد که بازدهی در روزهای مختلف هفته مستقل نبوده و در تقابل با تئوری گشت تصادفی قرار می‌گیرد. بدری و صادقی به این نتیجه رسیدند که اثر چهارشنبه (اثر آخر هفته) و نیز اثر روز یکشنبه در اکثر تخمین‌ها معنادار بوده است [۲۱].

اثر ماه‌های خاص سال در تقویم اسلامی (ماه‌های رمضان، محرم و صفر) : اثر ماه‌های خاص سال در تقویم اسلامی به عنوان نمونه‌هایی از بی‌قاعدگی‌های تقویمی بازار مورد توجه بسیاری از محققان قرار گرفته است که در آن شواهدی دال بر وجود بازده‌های واقعی و غیرعادی یافت می‌شود. در این ماه‌ها هم‌زمان با کاهش روند رو به رشد فعالیت‌های اقتصادی و عملیات سفته‌بازی مسلمانان، تعداد معاملات در بازارهای مالی کاهش می‌یابد. گرکز و اکرمی اقدام به بررسی اثر ماه رمضان بر بازده (واقعی و غیرعادی) کردند. نتایج آنان حاکی از آن بود که رابطه‌ی معناداری بین ماه رمضان و بازده واقعی وجود ندارد ولی با بازده غیرعادی سهام رابطه معناداری وجود دارد. معین‌الدین و عظیمی به بررسی تأثیر ماه محرم و صفر بر بازده سهام شرکت‌های پذیرفته شده در بورس پرداختند. نتایج حاکی از آن بود که بین ماه محرم و صفر و بازده سهام رابطه معناداری وجود دارد [۲۲,۲۳,۲۴,۲۵,۲۶].

اثر روزهای تعطیل با اثر قبل از روزهای تعطیل : مطالعات زیادی در کشورهای مختلف انجام شده که در آن شواهدی دال بر وجود بازده‌های غیرعادی بالا در روزهای قبل از تعطیلات رسمی یافت می‌شود. لیانو (۱۹۹۳) شواهدی از این امر را در بازارهای خارج از بورس نشان داد. بیشتر این مطالعات در بازارهای سهام آمریکا انجام شده و اکثر آن‌ها وجود چنین اثری، یعنی بازده‌های غیرعادی طی چند روز قبل از تعطیلات رسمی را تأیید کرده‌اند.

اثر آخر ماه و بین ماه : مطالعات اولیه مربوط به اثر بین ماه توسط آریل (۱۹۸۷) صورت پذیرفت. وی دریافت که به طور متوسط سهام شرکت‌ها در نیمه اول ماه (که از چند روز از آخر ماه قبلی شروع

می‌شود) بازده مثبت غیرعادی دارند. هم‌چنین بازده نیمه‌ی اول ماه به طور معناداری بیشتر از نیمه‌ی دوم ماه است. هم‌چنین آرساد و کوئز (۱۹۹۷) نیز چنین اثری را در بازده‌های ۶۰ روزه در بازار سهام لندن نشان داد.

اثر پایان سال (اثر ژانویه، ار دسامبر) : اثر ماه‌های سال مخصوصاً اثرات آخر سال مورد توجه بسیاری از محققان قرار گرفته است و اصلی‌ترین حوزه بررسی در مطالعات کاربردی بی‌قاعدگی‌های بازار را به خود اختصاص داده است. طی سال‌های گذشته مطالعات زیادی بیان‌گر آن بوده‌اند که بازدهی سهام، به خصوص سهام شرکت‌های کوچک در نخستین ماه سال میلادی نسبت به سایر ماه‌ها به طور معناداری بیشتر است. این پدیدیه که در بسیاری از بازارهای جهانی مورد بررسی قرار گرفته است، به اثر ژانویه (فروردین) معروف است. هم‌چنین آن‌چه در آخرین ماه سال در بازارهای مالی بسیاری از کشورها دیده شده و مورد بررسی قرار گرفته، وجود پدیده‌ایی در جهت عکس اثر ژانویه است که به اثر دسامبر (اسفند) معروف است. این اثر نیز بیان می‌دارد میانگین بازدهی سهام در آخرین ماه سال نسبت به سایر ماه‌ها کمتر است.

وجود اثرات تقویمی بر بازدهی بازار سهام، اقتصاددانان مالی را متحیر کرده است. بر اساس فرضیه بازار کارا، الگوی فصلی نباید دارای اثرات معنادار باشد. اثرات تقویمی باعث بازدهی‌هایی می‌شوند که متناسب با ریسک نیستند. به عبارت دیگر، وجود اثرات تقویمی شکل ضعیف فرضیه‌ی بازار کارا را خنثی می‌کند.

در ایران پژوهش‌های بسیاری در زمینه ارزیابی میزان کارایی بازار بورس اوراق بهادار تهران انجام گرفته که در ادامه به برخی از آنان اشاره شده است. با توجه به تحقیقات زهرا نصرالهی در سال ۱۳۷۱ بازار بورس اوراق بهادار تهران در سطح ضعیف کارا می‌باشد. پژوهش اسماعیل فدائی‌نژاد در سال ۱۳۷۳ دلالت بر عدم کارایی بورس اوراق بهادار تهران در سطح ضعیف دارد. حسنعلی سینایی در سال ۱۳۷۳در سطح نیمه‌قوی کارایی بازار بورس اوراق بهادار تهران را بررسی کرده و نتیجه تحقیق رد فرضیه کارایی بورس در سطح نیمه‌قوی بود و محمد نمازی و زکیه شوشتری در سال ۱۳۷۵ در مورد کارایی بازار در سطح ضعیف

تحقیق کرده‌اند که نتایج این آزمون نیز عدم کارایی در سطح ضعیف بورس اوراق بهادار تهران را نشان داد. نتایج حاصل از تحقیق حسنعلی سینایی در سال ۱۳۸۴ وجود روندی آشوبناک و غیرتصادفی در قیمت‌های سهام، به دلیل همبستگی بین قیمت‌های متوالی در سری زمانی نشان داد. علاوه بر آن طبق نتایج حاصل از پژوهش جواد عظیمی و همکاران در سال ۱۳۹۱، به دلیل وجود اثرات غیرعادی تقویمی از جمله اثر فروردین، اثر ماه محرم و صفر، اثر ماه رمضان و سایر اثرات ماهانه، الگوی خاصی در رفتار بازار وجود دارد که شناسایی و بهره‌گیری از روند این الگوها، سرمایه‌گذاران حرفه‌ایی را در جهت کسب بازده غیرعادی یاری می‌رساند، لذا می‌توان گفت بورس اوراق بهادار تهران حتی در سطح ضعیف یا نیمه قوی کارا نبوده و رفتار قیمتی سهام تابع الگوی تصادفی صرف نمی‌باشد و فرضیه گشت تصادفی قیمت‌ها رد شده و قیمت‌ها و بازده سهام در این بازار قابل پیش‌بینی می‌باشند.

۲-۱-۷ نظریه آشوب :

مطالعه در مورد تئوری آشوب[1] در حقیقت از سال ۱۹۶۵ توسط دانشمندی به نام ادوارد لورنز[2] از مطالعات هواشناسی شروع شد. این نظریه سپس در حیطه تمام علوم و مباحث تجربی، ریاضی، رفتاری، مدیریتی و اجتماعی وارد شده و اساس تغییرات بنیادی در علوم به ویژه هواشناسی، نجوم، مکانیک، فیزیک، ریاضی، زیست‌شناسی، اقتصاد و مدیریت را فراهم آورده است. نظریه آشوب به شاخه‌ای از ریاضیات و فیزیک گفته می‌شود که در ارتباط با سیستم‌هایی است که دینامیک آن‌ها در برابر تغییر مقادیر اولیه رفتار بسیار حساسی نشان می‌دهد به گونه‌ای است که رفتارهای آینده آن قابل پیش‌بینی نیست. تئوری آشوب را با کلید واژه‌هایی چون «نظم در بی نظمی» می‌شناسند. در واقع منظور از آشوب در ادبیات این تئوری، ناکارآمدی، به هم ریختگی و نظایر آن نیست. بلکه در اینجا همان گونه که هیلز۳

[1] : Chaos Theory
[2] : Edvard Lorenz
[3] : Hayles

توصیف می‌کند بی‌نظمی و آشوب، نوعی بی‌نظمی منظم یا نظم در بی‌نظمی است. بی‌نظم از آن رو که نتایج آن غیرقابل پیش‌بینی است و منظم بدان جهت که از نوعی قطعیت برخوردار است. بی‌نظمی در مفهوم عملی یک مفهوم ریاضی محسوب می‌شود که می‌توان آن را نوعی اتفاقی بودن همراه با قطعیت دانست. قطعیت آن به خاطر آن است که بی‌نظمی دلایل درونی دارد و به علت اختلالات خارجی رخ نمی‌دهد و اتفاقی بودن بدین علت است که رفتار بی‌نظمی، بی‌قاعده و غیرقابل پیش‌بینی است [۲۷].

ویژگی‌های تئوری آشوب :

۱ـ اثر پروانه‌ای[1]

دکتر لورنز در سال ۱۹۷۳ نتایج محاسبات دستگاه معادلات دیفرانسیل متشکل از سه معادله دیفرانسیل غیرخطی و معین مربوط به جابه جایی حرارتی جو را منتشر کرد. براساس مطالعات لورنز در محدوده معینی از عوامل معادلات، بدون دخالت عناصر تصادفی یا ورود اغتشاش خارجی نوع نوسانات نامنظم در پاسخ به سیستم بروز داده می‌شود. وی به این نتیجه شگرف رسید که یک تغییر جزیی در شرایط اولیه معادلات پیش بینی کننده وضع جوی منجر به نوسانات عظیمی در پاسخ سیستم و تغییرات شدید در نتایج حاصل از آن می‌شود. لورنز این خاصیت را اثر پروانه‌ای نامگذاری کرد. این مسئله، سنگ بنای تئوری آشوب است زیرا براساس نظریه آشوب اعتقاد بر آن است که در تمامی پدیده‌ها، نقاطی وجود دارد که تغییر اندک در آن‌ها باعث تغییرات عظیم خواهد شد، در این رابطه سیستم‌های اقتصادی، سیاسی، اجتماعی و سازمانی همچون سیستم‌های جوی از اثر پروانه‌ای برخوردارند. تحلیلگران باید با آگاهی از این نکته مهم به تحلیل و تنظیم مسائل مربوط بپردازند. اثر پروانه‌ای توجیهی عقلایی و کامل از رفتارها و تصمیم‌های مدیران خلاق و کارآفرین و موفق به دست می‌دهد که با یک حرکت مناسب و کم انرژی توانسته‌اند موجب تحول و دگرگونی‌های بزرگ و شگرفی در سازمان خود شوند [۲۸].

[1] : **Butterfly Effect**

۲ـ خودسازماندهی[1]

در محیط در حال تغییر امروزی سازمان سیستم‌های بی‌نظم در ارتباط با محیط اطرافشان مانند موجودات زنده عمل می‌کنند که برای رسیدن به موفقیت همواره باید خلاق و نوآور و در جست‌وجوی راه‌های جدیدی برای پیروزی و پیشرفت باشند. اما زمانی که سیستم به تعادل سازگار نزدیک می‌شود برای حفظ پویایی نیاز به تغییرات اساسی درونی دارد که این تغییرات به جای سازگاری و تطبیق با محیط، موجب سازگاری پویا می‌شود که نتیجه آن دگرگونی روابط پایدار بین افراد، الگوهای رفتاری، الگوهای کار، نگرش‌ها و طرز تلقی و فرهنگ هاست. برخی از دانشمندان معتقدند که آشفتگی، سازگاری و انطباق‌ها را در هم شکسته. براساس نظر مورگان، سیستم باید توان احساس و درک محیط و جذب اطلاعات از محیط را داشته باشد. دوم اینکه سیستم باید قادر به برقراری ارتباط بین این اطلاعات و عملیات باشد و سوم اینکه آگاهی از انحرافات و چهارم توانایی اجرای عملیات اصلاحی از خصوصیات اصلی این سیستم هاست [۲۹].

۳ـ خودمانایی[2]

در تئوری آشوب و معادلات عملیات آن نوعی شباهت بین اجزا و کل قابل تشخیص است. بدین گونه که هر جزیی از سیستم دارای ویژگی کل بوده و مشابه آن است. دنیس گابور در سال ۱۹۴۸ برای اولین بار هولوگرافی را بدین گونه بیان کرد: جزء خاصیت کل را دارد و مانند آن عمل می‌کند. سیستم توانایی یادگیری را دارد، سیستم دارای توانایی خود سازماندهی است، حتی اگر قسمت‌هایی از سیستم برداشته نشود سیستم به راحتی می‌تواند به فعالیت خود ادامه دهد.

از خاصیت خودمانایی می‌توان در سازماندهی سازمان‌های جدید بهره گرفت و سازمان‌هایی طراحی کرد که هر واحد آن به صورت خودکفا قادر به انجام وظایف سازمانی باشد[۳۰].

---

[1] : **Dynamic Adaptation**
[2] : **self- Similarity**

۴ـ جاذبه‌های عجیب

جاذبه‌های نقطه‌ای و دوره‌ای پایه‌های فیزیک نیوتنی کلاسیک است که بیانگر الگوی نظم و با ثبات در حرکت پدیده‌ها و روابط آنهاست مانند حرکت دادن یک مداد روی کاغذ حول محور خودش به شعاع یکسان که منجر به رسم دایره‌ای با شعاع مشخص خواهد شد .

غیرقابل پیش بینی بودن رفتار در جاذبه‌های عجیب تابع دو پدیده است: اولی مربوط به حساسیت نسبت به شرایط اولیه است که لورنز آن را اثر پروانه‌ای نامید دوم اینکه تغییرات شدید رفتارهای نامنظم و دگرگونی‌های غیرقابل پیش بینی حرکات، همه در درون خود نظمی نهفته دارند .

الگوی آشفتگی طرفداران بی‌شماری در میان محققان مالی پیدا کرده است. این پژوهشگران اذعان می‌کنند که رفتار قیمت‌ها به صورت غیرخطی است. حرکات قیمت سهام ناشی از عوامل بسیار زیادی است که در هر لحظه از زمان در آن‌ها تأثیر می‌گذارد. آنان در واقع معتقدند که به تعداد سرمایه‌گذاران، عامل تأثیرگذار بر قیمت سهام وجود دارد.

حسنعلی سینایی در سال ۱۳۸۴ طی پژوهشی تحت عنوان "پیش‌بینی شاخص بورس اوراق بهادار تهران با استفاده از شبکه‌های عصبی مصنوعی" وجود روندی آشوبناک و غیرتصادفی در قیمت‌های سهام را نشان داده است و حضور تئوری آشوب در این حوزه را تأیید نموده است.

۲-۲ بخش دوم : انواع روش‌های پیش‌بینی

موضوع شناخت و بررسی رفتار قیمت اوراق بهادار و ارزشیابی دارایی‌های مالی، از آغاز شکل‌گیری بازارهای سرمایه همواره مورد توجه محافل علمی و سرمایه‌گذاران بوده است و روش‌های سنتی و مدرن بسیاری برای نیل به این هدف ارائه شده است. از جمله مکاتب مالی در خصوص پیش‌بینی، رویکرد تکنیکی، بنیادی، رویکرد مبتنی بر نظریه‌های مدرن مالی و رویکرد بی‌نظمی و پویایی غیرخطی است[۱۵].

۲-۲-۱ روش‌های کلاسیک پیش‌بینی

قبل از وجود رایانه‌ها و استفاده از آن‌ها برای پیش‌بینی در حوزه‌های مالی، کار پیش‌بینی با روش‌های دیگری انجام می‌شده که در ادامه به برخی از آنان اشاره شده است.

۲-۲-۱-۱ رویکرد تجزیه و تحلیل بنیادین:

این رویکرد و مدل‌های مورد استفاده در آن از دهه ۱۹۳۰ مطرح بوده، اما عمدتاً بعد از جنگ جهانی دوم در قالب‌های نظری فراگیر، مورد توجه قرار گرفتند. در این‌گونه مدل‌ها اساساً به ارزش ذاتی سهم توجه می‌شود. این دسته از روش‌ها به دلیل تعیین ارزش ذاتی سهام به صورت علمی و با تکیه بر ابزارهای مختلف علمی از قبیل اقتصاد، آمار، اطلاعات مالی و ... مورد تأیید دانشگاهیان می‌باشد. مدل‌های مبتنی بر تحلیل بنیادی برای تعیین ارزش ذاتی سهام به صورت‌های مالی، سوابق تقسیم سود، سیاست‌های مدیریت، رشد فروش، توان مؤسسه در افزایش سودآوری و بسیاری عوامل دیگر توجه نموده، سپس ارزش ذاتی به‌دست آمده را با قیمت جاری سهام مقایسه می‌کنند. بر این اساس در مورد خرید، فروش و یا نگهداری آن تصمیم‌گیری می‌نمایند. در میان روش‌های بنیادین، آندسته نظریات که ارزش ذاتی سهام را در توان ایجاد درآمد شرکت جستجو می‌نمایند را مورد حمایت بیشتری قرار می‌دهند، زیرا با اصول

فرضیه بازار کارا نیز تطابق بیشتری دارد.

در این رویکرد برای تعیین ارزش ذاتی سهام از روشی تحت عنوان تجزیه و تحلیل اساسی، پایه‌ای یا بنیادین استفاده می‌شود که در واقع همان بکارگیری تمامی اطلاعات موجود و در دسترس است. در این روش وضعیت یک شرکت را با توجه به تمام ابعاد اصلی خرد و کلان، مورد توجه و ارزیابی قرار می‌دهند. به بیان دیگر، ابتدا اوضاع اقتصادی کشور را مد نظر قرار داده، وضعیت شاخص‌های با اهمیتی مثل GNP، نرخ تورم، نرخ ارز و به طور خلاصه سیاست‌های پولی و مالی دولت را بررسی می‌نمایند. سپس برای تعیین وضعیت صنعت مورد نظر، به نوع تولید، مصرف‌کنندگان و تقاضای عمومی، وجود کالاهای جانشین، حمایت‌های دولت، سیاست‌های قیمت‌گذاری و ثبات فروش توجه می‌کنند. در مرحله سوم به تجزیه و تحلیل شرکت پرداخته و پارامترهایی را از قبیل کیفیت مدیریت، کارایی عملیاتی، قابلیت سودآوری و بررسی ساختار مالی آن، مورد نظر قرار می‌دهند. سپس سهام شرکت‌ها به طور مجزا بر مبنای ریسک و بازده بررسی و انتخاب می‌شوند و نهایتاً سهام باقی‌مانده با پرتفویی از دارایی‌ها ترکیب می‌شوند [۳۱].

۲-۲-۱-۲ رویکرد تجزیه و تحلیل تکنیکال:

از اوایل قرن بیستم که به تدریج رفتار قیمت سهام و ارزش آن به شکلی علمی‌تر مورد توجه قرار گرفت، برخی از دست‌اندرکاران و شرکت‌های سرمایه‌گذاری از طریق تعقیب قیمت و روندهای خاص، الگوی تغییرات قیمت را بدست آورده و نتایج کارهای خود را مبنای تصمیمات سرمایه‌گذاری قرار می‌دادند. ترسیم رفتار قیمت، بررسی و تهیه نمودارها و مطالعات نوسانات و شناخت حساسیت‌های رفتار قیمت و پیش‌بینی آینده، هدف اصلی این گروه از صاحب‌نظران می‌باشد. هنوز نیز این تفکر مورد قبول بسیاری از سرمایه‌گذاران و مؤسسات می‌باشد. این گروه را تحلیل‌گران تکنیکی یا چارتیست می‌خوانند، زیرا از منحنی‌ها و نمودارها بسیار استفاده می‌نمایند. این تحلیل‌گران معتقدند که عوامل مؤثر بر عرضه و تقاضا

بی‌شمارند و هیچگاه نمی‌توان آن‌ها را به درستی و دقت شناسایی نمود، لذا بهترین شیوه کار را مطالعه حرکات گذشته و به دست آوردن الگوی تغییرات آینده می‌دانند. آنان عرضه و تقاضا را وابسته به عوامل بسیار زیادی دانسته و اعتقاد دارند که قیمت‌های گذشته، منعکس‌کننده آینده بوده و قیمت را تابع محض عرضه و تقاضا می‌دانند. آنان به دنبال تغییرات بلند مدت نیستند و می‌گویند باید از فرصت‌های کوتاه مدت حداکثر استفاده را نمود و سود آنی به‌دست آورد. این روش در حال حاضر نیز علاقمندان بسیار زیادی در میان تحلیل‌گران بازار دارد ولی به علت ضعف استدلال و توجیهات علمی، در مجامع دانشگاهی مورد قبول و حمایت نیست [۳۲].

۲-۲-۱-۳ رویکرد مبتنی بر نظریه‌های مدرن مالی:

پس از جنگ جهانی دوم، مجموع نظریه‌های جدیدی به جامعه مالی عرضه شد که عنوان نظریه‌های مدرن مالی را به خود اختصاص داد. دیدگاه‌های ارائه شده در این مطالعات، با بررسی‌های گذشته، تفاوتی اساسی داشت و در آن زمان مورد استفاده وسیع دست‌اندرکاران بازارهای سرمایه قرار گرفت. عمر نظریه‌های مدرن مالی در حدود نیم قرن است. ظرف این مدت این نظریه‌ها مبنای محاسبه ارزش دارایی‌های مالی و پیش‌بینی قیمت اوراق بهادار بوده و کوشیده‌اند وضعیت بازار سرمایه را توضیح دهند. سه حوزه به هم پیوسته فرضیه بازار کارا، نظریه پرتفوی و مدل قیمت‌گذاری دارایی‌های سرمایه‌ایی[1] که از حوزه‌های این رویکرد هستند، مورد تأیید محافل دانشگاهی‌اند. این نظریه‌ها بر بازارهای مالی جهان تأثیری عمیق گذاشته و پایه فکری مدیران مالی و سرمایه‌گذاری بوده و در سال‌های گذشته مرکز ثقل رد یا قبول سرمایه‌گذاری برای بسیاری از پروژه‌ها بوده‌اند.

برای اولین بار پیتر برن‌اشتاین[2] در کتاب انگاره‌های سرمایه پیچیده‌تر شدن فناوری و عدم امکان رسم

[1] : CAPM : Capital Asset Pricing Model
[2] : Peter Bernstein

نمودار و درک پیام نوسانات بازار را مطرح کرد و اظهار داشت که دیگر نمی‌توان مانند کولی‌ها از روی برگ چای، آینده قیمت‌ها را پیشگویی کرد. این دیدگاه و نتایج حاصل از برخی مطالعات موجب شد تا نظریه‌پردازان، به بررسی فرضیه‌هایی روی آورند که در آن قیمت سهام به طریق تصادفی حرکت می‌کرد. این مفهوم علمی، گشت تصادفی[1] نامیده شد و بتدریج در نشریه‌ها و کتاب‌های معتبر جهان جای خود را باز کرد و بر رویکرد مبتنی بر نظریه‌های مدرن مالی تأثیر گذاشت [۳۳].

مدل‌هایی که در ادامه به آن‌ها اشاره خواهد شد، زیرمجموعه‌ایی از مدل‌های اقتصادسنجی محسوب می‌شوند. ویژگی مدل‌های پیش‌بینی اقتصادسنجی استفاده از متغیرهای متعدد اقتصادی و اجتماعی است که به طور علی با متغیر وابسته ارتباط دارند. این مدل‌ها متشکل از یک یا چند معادله توصیف کننده رابطه بین متغیرهای مختلف هستند که سعی دارند تا رابطه میان متغیرهای وابسته و مستقل را با استفاده از یک یا چند معادله رگرسیون، توصیف نمایند. در حقیقت اقتصادسنجی روش‌هایی برای شناسایی و تخمین مدل‌هایی با چند مجهول را ایجاد می‌کند. این متدها به محقق اجازه می‌دهند که استنتاجی علی-معلولی در شرایطی غیر از شرایط آزمایشی کنترل شده ارائه دهد. مدل‌های اقتصادسنجی گستره‌ی وسیعی را به خود اختصاص داده‌اند که در نگاهی اجمالی می‌توان آن‌ها را با توجه به تعداد عواملی که با متغیر وابسته رابطه علی دارند، به دو دسته مدل‌های تک عاملی و مدل‌های چندعاملی تقسیم نمود. در ادامه به تفصیل به توضیح این مدل‌ها پرداخته شده است.

۲-۲-۱-۳-۱ مدل قیمت گذاری دارایی های سرمایه ایی (CAPM)

این مدل که یکی از معروف‌ترین مدل‌های شناسایی عوامل تأثیرگزار بر بازده سهام و پیش‌بینی آن محسوب می‌شود، توسط شارپ (۱۹۶۴)، لینتنر (۱۹۶۵) و موسین (۱۹۶۶) بر اساس چارچوب مارکویتز ارائه شده است. این مدل فرض می‌کند که سرمایه‌گذاران از منطق ماکویتز در تشکیل پرتفلیو استفاده

[1] : Random Walk

می‌کنند. فرض دیگر آن وجود یک دارایی (دارایی بدون ریسک) است که بازدهی مشخصی دارد. این فرض در ارزش‌گذاری دارایی‌ها با نرخ کاهش مناسب که در هر مدل ارزش‌گذاری به کار می‌رود، بسیار راه‌گشا بوده و اهمیت فراوانی دارد. به بیان دیگر اگر بتوان نرخ بازدهی را که از یک سرمایه گذاری به دست می آید، برآورد کرد، می توان این نرخ بازدهی برآورد شده را با نرخ بازدهی مورد انتظار آن که بر اساس مدل CAPM حاصل می‌شود را مقایسه کرده و تعیین کرد که آیا این دارایی کمتر، بیشتر و یا به درستی قیمت‌گذاری شده است [۳۴].

پژوهش‌های مقطعی اولیه در مورد بازده سهام همچون نیکلسون (۱۹۶۰) به دلیل نمونه‌های کوچک استفاده شده برای انجام آزمون‌های تجربی مورد توجه کافی قرار نگرفت، تا اینکه پایگاه داده‌های CRSP[1] و Compustat[2] به وجود آمد و پژوهشگران توانستند نمونه‌هایی به اندازه کافی بزرگ (و با کیفیت کافی) برای نتایج قابل اطمینان ایجاد کنند. در پی آن پس از چند سال ارائه مدل CAPM هیچ راه قابل اطمینانی برای آزمون پیش‌بینی‌های این مدل در مورد متغیرهایی همچون نسبت ارزش بازار به ارزش دفتری و نسبت قیمت به سود وجود نداشت.

در ادامه‌ی پژوهش‌های مهم انجام شده در رابطه با نقض مدل CAPM متغیرهای مورد بررسی، به تفکیک شرح داده شده‌اند. این موارد در ادبیات مالی با عنوان بی‌قاعدگی‌های بازار[3] شناخته می‌شوند. در حقیقت بی‌قاعدگی‌های بازار نتایج پژوهش‌های تجربی هستند که با تئوری‌های مدون قیمت‌گذاری دارایی‌ها هم‌خوانی ندارند. این بی‌قاعدگی‌ها نشان‌دهنده ناکارآمدی بازار (فرصت‌های سودآوری) یا کامل نبودن مدل قیمت‌گذاری دارایی مورد استفاده می‌باشد.

۱. نسبت سود به قیمت (E/P):

یکی از مطالعات اولیه که پیش‌بینی‌های CAPM را رد کرد، پژوهش باسو (۱۹۷۷) بود. باسو با

[1] : CRSP : Center for Research in Security Prices
[2] : Compustat : Computer Statistics
[3] : Market Anomalies

استفاده از یک دوره زمانی از آوریل ۱۹۵۷ تا مارس ۱۹۷۱، نشان داد که سهامی که نسبت سود به قیمت بالایی داشتند (یا نسبت P/E پایینی داشتند) به صورت معنی‌داری بازده‌هایی بیش از سهام با نسبت سود به قیمت پایین تحصیل کردند. پژوهش بعدی که توسط جف، کیم و وسترفیلد (۱۹۸۹) انجام شد، علاوه بر تأیید این نتیجه نشان داد که بر خلاف آن چه که توسط برخی پژوهشگران ادعا می‌شود، اثر E/P نقض مستقیم مدل CAPM (که باور دارد تنها ریسک سیستماتیک باید مبنا قرار گیرد) می‌باشد.

۲. اندازه‌ی شرکت :

بنز (۱۹۸۱) نشان داد که سهام شرکت‌های با ارزش بازار پایین، بازدهی متوسط بالاتری از سهام شرکت‌های با ارزش بازار بالا دارند. سایر پژوهشگران نظیر باسو (۱۹۸۳) نشان دادند که اثر اندازه‌ی متمایز از اثر E/P می‌باشد. شرکت‌های کوچک بازدهی‌های بالاتری دارند.

۳. برگشت معکوس بازده در بلند مدت[1]

دبوند و تالر (۱۹۸۵) بازنده‌گان را به عنوان سهامی که طی سه تا پنج سال گذشته بازدهی‌های پایینی داشته‌اند و برندگان را به عنوان سهامی که بازدهی‌های بالایی را در خلال یک دوره‌ی مشابه داشته‌اند، تعریف کردند. نتیجه اصلی پژوهش این دو نشان داد که در خلال سه تا پنج سال آینده، بازنده‌گان بازدهی متوسط بالاتری را نسبت به برندگان داشته‌اند. این حرکت بازدهی‌ها نسبت به معکوس شدن طی افق‌های بلند مدت (یعنی برنده شدن بازنده‌گان) هنوز هم یکی از موارد نقض CAPM می‌باشد. کپرا، لاکونیشوک و ریتر (۱۹۹۲) نشان دادند که تفاوت‌های ریسک سیستماتیک (β) در این زمینه برای توجیه مدل CAPM کافی نمی‌باشد.

۴. نسبت ارزش دفتری به ارزش بازار (B/M):

رزنبرگ، رید و لانستین (۱۹۸۵) نشان دادند که سهم‌های با نسبت ارزش دفتری به ارزش بازاری بالا

[1] : Long – Term Return Reversal

به صورت معناداری بازدهی‌های بالاتری نسبت به سهام با نسبت ارزش دفتری به ارزش بازار پایین دارند. چان، هامائو و لاکونیشوک (۱۹۹۱) به نتایج مشابهی در بازار ژاپن دست یافتند.

۵. اهرم مالی[1] :

بانداری (۱۹۹۸) دریافت که شرکت‌های با اهرم مالی بالا (نسبت "بدهی به سرمایه" بالا) از شرکت‌های با اهرم مالی پایین در خلال دوره‌ی ۱۹۷۹-۱۹۸۴ به طور متوسط بازده بالاتری داشتند.

به بیان بهتر می‌توان گفت که آزمون‌های تجربی اولیه CAPM پیش‌بینی محوری آن را مبنی بر وجود رابطه خطی مثبت بین ریسک سیستماتیک (بتا) و بازده سهام، مورد تأیید قرار دادند، با این وجود نتایج مطالعات اخیر حکایت از این دارد که ضریب بتا به عنوان شاخص ریسک سیستماتیک، توان تشریح اختلاف میانگین بازده سهام را ندارد و غیر از بتا، سایر متغیرها در چارچوب مدل CAPM محلی از اعراب ندارند. از آن جمله می‌توان به اندازه شرکت، نسبت ارزش دفتری به ارزش بازار، نسبت سود به قیمت و اهرم مالی که در تبیین اختلاف بازده سهام نقش مؤثری ایفا می‌کنند، اشاره کرد. در واقع ریسک‌های بی‌شماری در ارتباط با شرکت وجود دارد. برخی از این ریسک‌ها شامل ریسک بازار، ریسک ورشکستگی و ریسک نقدینگی است، در حالی که در مدل CAPM تنها از یک عامل جهت تشریح مجموع ریسک‌ها استفاده می‌شود. از نظر منطقی، به نظر می‌رسد مدلی با عوامل گوناگون، ممکن است قدرت توصیف‌کنندگی و پیش‌بینی‌کنندگی بیشتری را فراهم کند. به عبارت دیگر، عوامل اضافه، اجازه می‌دهند ریسک‌هایی توزیع شوند که شرکت با آن مواجه است. عامل ریسک بازار صرفا ابعاد گوناگون ریسک را تجزیه می‌کند و قادر به بیان، تشریح و توصیف اثر انواع ریسک بر بازده نمی‌باشد. طی سال‌های ۱۹۸۰ تا ۱۹۹۰ انحرافات و ناهنجاری‌های مدل CAPM آشکار گردید. به عقیده پژوهشگران، این ناهنجاری‌ها به عنوان چالشی بر اعتبار CAPM در توانایی تشریح بازده مورد انتظار توسط عامل ریسک سیستماتیک (بتا) مطرح می شود [۳۵].

[1] : Financial Leverage

بر خلاف همه‌ی این شواهد منفی، CAPM تا اوایل دهه ۱۹۹۰ هنوز هم به عنوان پیش فرض توسط اغلب تحلیل‌گران مالی به کار گرفته می‌شد.

۲-۲-۱-۳-۲ مدل فاما و فرنچ :

یوجین فاما[1] که رساله دکتری خود را در سال ۱۹۶۵ به بررسی گشت تصادفی حرکت قیمت‌های سهام با استفاده از مفروضات معین اختصاص داده بود، در سال ۱۹۹۲، مقاله‌ای منتشر کرد که بیشتر پژوهش‌های تجربی قبلی را یک جا جمع‌آوری کرده بود. فاما و فرنچ (۱۹۹۲) متغیرهای اندازه، اهرم مالی، E/P ، ارزش دفتری به ارزش بازار و بتا را در یک پژوهش مقطعی[2] آزمون کردند. آن‌ها قدرت توضیح دهندگی اندازه، اهرم، E/P ، نسبت ارزش دفتری به ارزش بازار و ریسک سیستماتیک را در رگرسیون مقطعی طی دوره ۱۹۹۰-۱۹۶۳ مقایسه کردند. نتایج نشان داد که اندازه شرکت و نسبت ارزش دفتری به ارزش بازار (ارزش شرکت) متغیرهایی هستند که قوی‌ترین رابطه را با بازده دارند و هنگامی که این دو متغیر در رگرسیون وارد می‌شوند، قدرت توضیح دهندگی سایر متغیرها ناپدید می‌گردد و بازده‌های متوسط مقطعی می‌توانند به صورت مفیدی توسط این دو متغیر توضیح داده شوند [۳۶].

نتایج فاما و فرنچ در سال ۱۹۹۲ این دیدگاه را که مدل تک عاملی CAPM روشی است که توسط آن اوراق بهادار به صورت واقعی قیمت‌گذاری می‌شوند، را خدشه‌دار ساخت. مدلی که بیش از هر مدل دیگر در دانشکده‌های مالی آموزش داده شده بود، به نظر می‌رسید که دیگر جواب‌گو نباشد. بر اساس این نتایج آن‌ها در سال ۱۹۹۳ مدل سه عاملی خود را ارائه کردند و نشان دادند که اندازه و نسبت ارزش دفتری به ارزش بازار دو عامل ریسک هستند که در CAPM نادیده گرفته شده اند. به عقیده آن‌ها، نسبت‌های قیمتی مختلف حاوی اطلاعات بسیار مشابهی در مورد بازده مورد انتظار هستند. آن‌ها به نتیجه مشابهی با استفاده از رویکرد رگرسیون سری زمانی در ارتباط با پرتفلیوهای سهامی دست یافتند که بر اساس قیمت مرتب شده بودند.

[1] : Eyogin Fama
[2] : Cross Sectional Research

عامل اندازه شرکت‌ها :

فاما و فرنچ به منظور تعیین اثر اندازه شرکت‌ها به ایجاد پرتفوهایی بر مبنای دو عامل اندازه و بتای سهام اقدام کردند. نتایج نشان داد که برای شرکت‌هایی که از نظر اندازه در یک سطح قرار گرفته‌اند، رابطه معناداری بین عامل بتا و بازده سهام وجود ندارد و در واقع این عامل اندازه است که توانایی تبیین نوسانات بازده را در این شرکت‌ها دارا می‌باشد.

عامل ارزش شرکت‌ها :

پژوهش‌های بسیاری از نسبت مبلغ دفتری به ارزش بازار حقوق صاحبان سهام برای ارزش‌گذاری این متغیر به عنوان یک عامل خطر برای توضیح نوسانات بازده سهام بهره برده‌اند. در این زمینه گریفین و لمون ثابت کرده‌اند که اگر بازده یک سهم در قبال هر واحد تغییر در این نسبت حساسیت بیشتری از خود بروز داد، ریسک خرید آن سهام بیشتر است. از نظر فاما و فرنچ این نسبت در مقایسه با متغیرهای مشابه پارامتر ممتازتری برای توضیح و پیش‌بینی نوسانات بازده سهام به حساب می‌آید..

مدل سه عاملی فاما و فرنچ نیز در معرض انتقادهای فراوانی قرار گرفت که این انتقادها در چارچوب سه موضوع محوری داده کاوی، اشتباه انتخاب و روش برآورد ریسک سیستماتیک مطرح شد. مخالفت هایی که با پژوهش فاما و فرنچ شد و حمایت هایی که از آن صورت گرفت، یکی از جالب ترین شاخه های ادبیات علوم مالی می باشد. به ندرت یک مقوله دانشگاهی را می توان یافت که به این اندازه در دنیای واقعی کاربرد داشته باشد.

هدف اصلی و اولیه تجزیه و تحلیل اوراق بهادار نسبت به تغییرات آنها، یافتن تعداد عوامل مؤثر بر نرخ‌های بازدهی اوراق می‌باشد. به بیان رسمی به چنین رابطه‌ایی اصطلاحاً مدل عاملی نرخ‌های بازدهی اوراق بهادار گفته می‌شود.

این مدل‌ها به دو گروه مدل‌های تک‌عاملی و مدل‌های چندعاملی تقسیم می‌شوند. مفهوم اساسی در مدل

تک عاملی این است که تمامی اوراق بهادار از نوسانات عمومی بازار تأثیر می‌پذیرند، زیرا نیروهای اقتصادی مشابهی در آینده اکثر شرکت‌ها را تحت تأثیر قرار می‌دهد. از مدل‌های تک عاملی می‌توان به مدل CAPM و مدل بازار اشاره کرد. ولیکن هدف اصلی در مدل‌های چند عاملی یافتن برخی از تأثیرات غیر بازاری است که منجر به حرکت توأم سهام با یکدیگر می‌شود چرا که با مطالعه آزمون‌های تجربی مدل تک عاملی CAPM، مشاهده می‌شود که عوامل دیگری در توضیح بازده سهام نقش دارند. این موضوع منجر به شکل‌گیری مدل‌های چند عاملی شده است.

۲-۲-۱-۳-۳ سایر مدل‌های چندعاملی :

مدل‌های چندعاملی فرض می‌کنند که بازده اوراق بهادار به حرکات عوامل و یا شاخص‌های متعددی حساس است، در حالی که در مدل بازار فرض بر این است که اوراق بهادار صرفاً از شاخص بازار تبعیت می‌کند. به منظور برآورد صحیح‌تر بازده مورد انتظار، واریانس و همبستگی اوراق بهادار، مدل‌های چندعاملی توانایی بالاتری از مدل‌های بازار دارند. علت این امر آن است که بازده واقعی اوراق بهادار به عواملی بیش از صرف شاخص بازار حساس می‌باشد. به عبارت دیگر، در اقتصاد واقعی بیش از یک عامل تأثیرگذار بر بازده اوراق بهادار وجود دارد.

به‌عنوان یک فرآیند ایجاد بازده، مدل‌های چندعاملی سعی دارند تا نیروهای عمده اقتصادی که به طور سیستماتیک قیمت‌های کلیه اوراق بهادار را تغییر می‌دهند را کشف کنند. البته سرمایه‌گذاران نیز مدل عاملی را به طور ضمنی و یا صراحتاً در تصمیم‌گیری‌های خود به کار می‌گیرند. برای برآورد مدل‌های چندعاملی روش‌های مختلفی مورد استفاده قرار می‌گیرد که این روش‌ها را می‌توان در بدو امر به سه دسته طبقه‌بندی کرد: ۱) روش سری زمانی ۲) روش مقطعی ۳) روش تجزیه و تحلیل عامل[1]

در روش سری زمانی فرض بر آن است که تحلیلگر عوامل تأثیرگذار بر بازده سهام را پیشاپیش میداند. شناخت عوامل مرتبط معمولاً از تجزیه و تحلیل‌های عوامل اقتصادی بر کل شرکت از جمله عوامل خرد و

[1] : Factor Analysis

کلان اقتصادی، ساختار صنعت و تجزیه و تحلیل اساسی اوراق بهادار بدست می‌آیند.

در روش مقطعی ابتدا حساسیت اوراق بهادار به عوامل معین تخمین زده می‌شود و سپس در دوره زمانی خاصی ارزش عوامل مختلف بر مبنای بازده اوراق بهادار و حساسیت آن‌ها به این عوامل برآورد می‌شود. این فرآیند طی چندین دوره تکرار شده به طوری که برآوردی از انحراف معیار عوامل و همبستگی آن‌ها بدست می‌آید.

در روش تجزیه و تحلیل عامل نه ارزش عوامل و نه حساسیت این عوامل مشخص است و از تکنیک آماری تجزیه و تحلیل عامل برای استخراج تعداد و حساسیت عوامل متعدد استفاده می‌شود. در هر صورت مدل برازش شده در مدل چندعاملی به شکل رابطه (۲-۷) می‌باشد:

رابطه (۲-۷) $$R_{it} = a_i + b_{i1}F_{1t} + b_{i2}F_{2t} + ... + b_{in}F_{nt} + e_{it}$$

در مدل فوق R_{it} بازده سهام i در زمان t و a_i، مقدار ثابت و b_i برای هر سهم عددی ثابت و نشان‌دهنده حساسیت هر عامل، F عامل مشترک برای کلیه سهام و e_i جز تصادفی برای هر سهم می‌باشد.

یکی از این مدل‌های چند عاملی مدل کارهارت می‌باشد. این مدل شکلی از مدل سه عامله فاما و فرنچ است که کارهارت در سال ۱۹۹۷ این مدل چهار متغیره را با اضافه کردن یک متغیر جدید تحت عنوان عامل شتاب ابداع کرد. مدل کارهارت بعدها توسط اوسط سو و چن آزمون شد و نشان داد که این مدل چهار متغیره قدرت بیشتری در پیش بینی بازده نسبت به مدل سه متغیره فاما و فرنچ و مدل تک متغیره CAPM دارد

در میان مدل‌های چندعاملی، تکنیک‌های پیش‌بینی سری زمانی بیش از سایر روش‌ها توجه محققان را به خود جلب کرده است. تجزیه و تحلیل سری‌های زمانی با تئوری ارزش‌گذاری دارایی‌ها در طول زمان گره خورده است. تفاوت مهمی که سری‌های زمانی مالی را از سایر سری‌های زمانی جدا می‌کند این است که تئوری‌های مالی و بنیان سری‌های زمانی آن، با نوعی عدم اطمینان همراه است و به همین علت است که تئوری‌ها و روش‌های آماری نقش اساسی‌ای در تجزیه و تحلیل سری‌های زمانی مالی ایفا می‌کنند.

سری‌های زمانی مالی انواع مختلف دارند و به شیوه‌های مختلفی قابل تقسیم‌بندی می‌باشند که مهم‌ترین و پرکاربردترین آن‌ها عبارت‌اند از:

الف: سری‌های زمانی خطی[1]

این دسته از سری‌های زمانی که بنیان و ریشه سایر مدل‌ها را تشکیل می‌دهند، چارچوب ساده‌ایی به منظور مطالعه ساختارهای پویا مهیا می‌سازند. مطالعه این مدل‌ها شامل وابستگی‌های ایستا و پویا، تابع خودهمبستگی، مدل‌سازی و پیش‌بینی می‌شوند [۳۷]. برخی از این مدل‌ها به صورت زیر تقسیم‌بندی می‌شوند:

۱: مدل‌های خودتوضیح رگرسیون ساده (AR)[2]

۲: مدل‌های میانگین متحرک ساده (MA)[3]

۳: مدل‌های تلفیقی خودتوضیح میانگین متحرک ساده (ARMA)[4]

۴: مدل‌های فصلی[5]

۵: مدل‌های پویای تک ریشه‌ایی [6]

[1] : Linear Time Series
[2] : Simple Autoregressive (AR) Models
[3] : Simple Moving-Average (MA) Models
[4] : Mixed Autoregressive Moving-Average (ARMA) Models
[5] : Seasonal Models
[6] : Unit-Root Nonstationary

ب: سری‌های زمانی ناهموار شرطی[1]

این دسته از تکنیک‌ها به مدل‌سازی نوسانات و تغییرپذیری [2] بازده دارایی می‌پردازد. در این‌جا منظور از تغییرپذیری و نوسان، انحراف معیار شرطی بازده دارایی مورد بررسی است که در مدیریت ریسک دارایی‌ها کاربرد فراوان دارد و موجب بهبود تخمین پارامتر و افزایش دقت پیش‌بینی می‌شود [۳۸]. مهم‌ترین مدل‌های معادلات تک متغیره تغییرپذیری عبارت‌اند از:

۱: مدل رگرسیون ناهموار خودتوضیح شرطی (ARCH)[3]

۲: مدل رگرسیون ناهموار خودتوضیح شرطی تعمیم‌یافته (GARCH)[4]

۳: مدل رگرسیون ناهموار خودتوضیح شرطی تعمیم‌یافته نمایی (EGARCH)[5]

۴: مدل رگرسیون ناهموار خودتوضیح شرطی تعمیم‌یافته آستانه‌ایی (TGARCH)[6]

۵: مدل رگرسیون ناهموار خودتوضیح شرطی میانگین متحرک (CHARMA)[7]

ج: سری‌های زمانی غیرخطی[8]

این دسته از مدل‌ها که با پیشرفت تکنولوژی روند رو به رشدی داشته‌اند، توسط تست‌های پارامتری و غیرپارامتری آماری از سری‌های زمانی خطی قابل تمیز می‌باشند. مفاهیم بنیادی این دسته از سری‌های زمانی استفاده از مدل‌های شبیه‌سازی به منظور توصیف تکامل توزیع شرطی متغیر مربوطه و یا خصوصیات غیرخطی آن‌ها می‌باشد [۳۹,۴۰]. مهم‌ترین این مدل‌ها عبارت‌اند از:

---

[1] : Conditional Heteroscedastic Models
[2] : Volatility
[3] : Autoregressive Conditional Heteroscedastic (ARCH) Model
[4] : Generalized ARCH (GARCH) Model
[5] : Exponential GARCH (EGARCH) Model
[6] : Threshold GARCH (TGARCH) Model
[7] : Conditional Heteroscedastic Autoregressive Moving Average (CHARMA)
[8] : Non-Linear Time Series

۱: مدل دوسویه[۱]

۲: مدل رگرسیون خودتوضیح آستانه‌ایی (TAR)[۲]

۳: مدل رگرسیون خودتوضیح انتقال هموار (STAR)[۳]

۴: مدل راه‌گزینی مارکو[۴]

۵: مدل‌های غیرپارامتری[۵] مثل رگرسیون کرنل[۶]

۲-۲-۲ روش‌های مدرن پیش‌بینی

به دنبال تلاش‌های دانشمندان علوم ریاضی و سیستم‌های پویا روش‌های جدیدی برای پیش‌بینی فعالیت‌های مالی ایجاد شده است. کاربرد مدل‌های غیرخطی و هم‌چنین تکنیک‌های پیشرفته در مدت زمان کمی توانسته است، جایگاه خود را در علوم مختلف، به ویژه اقتصاد باز کند. متخصصین سیستم‌های غیرخطی، سعی در توضیح رفتار بازارهای مالی و پیش‌بینی آن از طریق روش‌های پیشرفته غیرخطی کرده‌اند. این روش‌ها عبارت‌اند از نظریه آشوب و روش‌های هوش مصنوعی.

از آن‌جایی که پیش‌بینی‌های بازارهای مالی، غیرخطی، پیچیده و غیرمتریک هستند و از مدل بی‌نظمی در طبیعت تبعیت می‌کنند، پیش‌بینی آن با روش‌ها و متدهای متداول مانند سری زمانی و رگرسیون ، به سختی می‌توان به پاسخ صحیح دست یافت. در این‌گونه مسائل مدل‌های شبیه‌سازی، داده‌کاوی، سیستم‌های خبره و روش شبکه‌های عصبی برای رفتارهای ناشناخته متغیرهای دارای نوسان بالا، ایده‌ال به نظر می‌رسد.

---

[1] : Bilinear Model
[2] : Threshold Autoregressive
[3] : Smooth Transition AR (STAR) Model
[4] : Markov Switching Model
[5] : Nonparametric Methods
[6] : Kernel Regression

۲-۲-۲-۱ رویکرد بی‌نظمی و پویایی غیرخطی:

از اواسط دهه ۱۹۷۰و به ویژه از سال ۱۹۹۰ کوشش‌های جدید و گسترده‌ای در زمینه پیش‌بینی قیمت سهام، با استفاده از روش‌های جدید ریاضی، سری‌های زمانی طولانی و ابزار پیشرفته آغاز گردید، که منجر به ظهور دیدگاه بی‌نظمی و پویایی‌های غیرخطی شد. پس از آن مطالعات بسیاری بر روی اطلاعات قیمت و شاخص سهام در کشورهای دارای بازارهای مالی پیشرفته انجام شد تا وجود ساختاری معین در اطلاعات قیمت سهام نشان داده شود که از این طریق فرضیه گشت تصادفی را نقض کنند.

اندیشمندان در این چارچوب، در جست و جوی روش‌های علمی جدیدتری هستند و معتقدند، نظریه بی‌نظمی و پویایی غیرخطی می‌تواند مقدمه ارائه نظریه‌های علمی تازه‌ایی در مدیریت مالی باشد. رنگ باختن نظریه‌های مدرن مالی به همراه دو پدیده جهانی شدن بازارهای مالی و وجود قدرت‌های تکنولوژیک در سطح جهان، موجب شده است تا نظریه‌های مدرن مالی در بازارهای مختلف به سادگی مورد آزمون قرار گیرند و ناتوانی آن‌ها در تبیین اقتصاد برای صاحبان اندیشه، روشن شود. بنابراین، تناقض‌های آشکار میان نظریه‌های موجود و دنیای واقعی، موجب شده است تا دانشگاهیان و اهل نظر، نسبت به اعتبار نظریه‌های مدرن مالی موجود، تردید کنند و فرضیه بازار کارا و به ویژه مدل یک متغیره شارپ[۱] را به دیده تردید بنگرند.

گروهی از صاحب‌نظران و اندیشمندان قدم فراتر گذاشته و به پدیده آشفتگی[۲] توجه نمودند. آن‌ها با استفاده از فنون جدید ریاضی، بازار را سیستمی پیچیده و در حال تحول می‌دانند و معتقدند، اگر زمینه و جوهره بازار به درستی درک شود، هر قفلی در آن گشودنی است. این گروه مانند ریاضی‌دانان و فیزیک‌دانان، اعتقاد دارند چنانچه هر پدیده‌ای به درستی مشاهده و بررسی شود، وقایع به ظاهر تصادفی آن تبیین‌پذیر است و اگر درخور پیش‌بینی نباشد، حداقل قابل درک خواهد بود.

---

[1] : Sharp
[2] : Chaos Theory

طرفداران الگوی آشفتگی اذعان می‌کنند که رفتار قیمت‌ها به صورت غیرخطی است. حرکات قیمت سهام ناشی از عوامل بسیار زیادی است که در هر لحظه از زمان در آن‌ها تأثیر می‌گذارد. آنان در واقع معتقدند که به تعداد سرمایه‌گذاران، عامل تأثیرگذار بر قیمت سهام وجود دارد. در این راستا به دلیل وجود روابط غیرخطی پیچیده با متغیرهای متعدد، توانایی و قابلیت مدل‌های رگرسیون خطی به دلیل زمان‌بر بودن آن‌ها در پردازش اطلاعات کاهش می‌یابد و لذا باید از یک سری روابطی مبتنی بر محاسبات نرم[1] استفاده شود.

۲-۲-۳ پیشینه تحقیق :

اصلی‌ترین عاملی که هر سرمایه‌گذار در تصمیمات خود مورد توجه خاص قرار می‌دهد، نرخ بازده است. یعنی سرمایه‌گذاران به دنبال پربازده‌ترین فرصت‌ها برای سرمایه‌گذاری مازاد منابع خود در بازارهای سرمایه هستند. از طرفی در زمینه پیش‌بینی بازده سهام که یکی از موضوع‌های مورد علاقه سرمایه‌گذاران و محققان مالی است تا کنون تلاش‌های زیادی برای ارائه مدلی که بتواند بازده سهام را به نحو درخور اتکایی پیش‌بینی کند، انجام شده است[۱۴].

پژوهش‌های اخیر توجه بیشتری به بازده سهام و پویایی دارند که به وسیله رفتار غیرخطی توصیف می‌شود چرا که بازارهای سهام رفتاری آشوب‌گونه دارند و نه تصادفی. حرکت بازده و قیمت سهام در طول زمان از عوامل مختلفی تأثیر می‌پذیرد و اینگونه رفتار ایستا نبوده و حرکتی پویا را نشان میدهد، لذا مدل‌های خطی و استاتیک قادر به تبیین رفتار چنین سیستم‌هایی نیستند و به همین دلیل گرایش به استفاده از مدل‌های غیرخطی روز به روز افزایش یافته است.

از میان مدل‌های غیرخطی، مدل‌های سری زمانی غیرخطی[2] توجه بسیاری از محققان را به خود جلب

[1] : Soft Computing
[2] : Nonlinear Time Series Models

کرده است و پژوهش های بی‌شماری در این زمینه انجام شده است. به منظور پیش‌بینی بازده سهام از مدل‌های خود توضیح آستانه‌ایی[1]، مدل خود توضیح انتقال هموار[2]، مدل خود توضیح انتقال هموار لجستیک[3] و .... به وفور استفاده شده است. دسته دیگر مدل‌های سری زمانی غیرخطی که پژوهش‌های بسیاری را به خود اختصاص داده‌اند، مدل شبکه عصبی مصنوعی است. این مدل به دلیل ویژگی‌های غیرخطی و ناپارامتری بودن و برخورداری از ساختاری داده محور، در تخمین مقادیر موفق‌تر عمل کرده است. تکنیک داده‌کاوی راف ست با توجه به مدت زمان کوتاهی که در تجزیه و تحلیل مسائل مالی کاربردی شده است، توجه بسیاری از محققان را به خود جلب کرده و عملکرد مناسبی با دقت بالا از خود نشان داده است. در ادامه به برخی از این پژوهش‌ها در داخل و خارج از کشور اشاره شده است.

۲-۲-۳-۱ پژوهش‌های انجام گرفته در ایران

در ایران پژوهش‌های بسیاری در زمینه پیش‌بینی قیمت سهام، بازده سهام و شاخص اوراق بهادار با استفاده از مدل‌های غیرخطی سری زمانی، شبکه‌های عصبی مصنوعی و تئوری فازی انجام شده است، در حالی که تاکنون پژوهشی در زمینه استخراج قواعد و قوانین حاکم بر بازده سهام با استفاده از تکنیک راف ست ارائه نشده است، بنابراین در این قسمت به بیان پیشینه پژوهش‌های مشابه که با سایر تکنیک‌ها انجام شده، پرداخته شده است که از این جمله می‌توان به پژوهش‌های زیر اشاره کرد.

علی رحمانی و فرشته سعیدی (۱۳۸۷) در پژوهش خود به منظور پیش‌بینی بازده سهام با استفاده از اطلاعات حسابداری و صورت‌های مالی از مدل‌های لاجیت استفاده کردند و نتایج آن را با روش رگرسیون مقایسه نمودند. نتایج این پژوهش نشان داد که عملکرد دو مدل تفاوت معنی‌داری ندارند.

ابراهیم عباسی و سحر باقری (۱۳۹۰) بازده سهام را با استفاده از مدل‌های غیرخطی خودتوضیح آستانه‌ای

[1] : TAR : Threshold Autoregressive
[2] : STAR : Smooth Transition Autoregressive
[3] : LSTAR : Logistic Smooth Transition Autoregressive

(TAR) و خودتوضیح انتقال هموار لجستیک (LSTAR) پیش‌بینی نموده و با مدل خطی ARMA مقایسه کردند. نتایج پژوهش آنان دقت پیش‌بینی بیشتر مدل‌های غیرخطی نسبت به مدل ARMA را نشان داد.

نظیفی نایینی، فتاحی و صمدی (۱۳۹۱) با استفاده از مدل انتقالی گارچ مارکف به پیش‌بینی نوسانات بازار بورس پرداختند. نتایج پژوهش آن‌ها حاکی از آن است که برای پیش‌بینی نوسانات بازار سهام ایران، عملکرد مدل‌های SW-GARCH با توزیع خطای t و با درجه آزادی متغیر بین دو رژیم، بسیار بهتر از مدل‌های گارچ معمولی است. حتی در برازش و بررسی‌های داخل نمونه‌ایی نیز این نوع از مدل‌های انتقالی مارکف، رتبه اول را در زمینه قدرت به خود اختصاص دادند.

زهرا پورزمانی و علی بشیری (۱۳۹۲) طی پژوهش خود در زمینه ارزیابی مدل کارهارت برای پیش‌بینی بازده مورد انتظار از مدل‌های رگرسیون چند متغیره و آزمون زوجی استفاده کردند.

رضا راعی و کاظم چاوشی (۱۳۸۲) در پژوهش خود به پیش‌بینی رفتار بازده سهام شرکت توسعه صنعتی بهشهر به وسیله مدل خطی عاملی (رگرسیون خطی چندمتغیره) و شبکه‌های عصبی مصنوعی با معماری MLP و الگوریتم پس‌انتشار خطا پرداختند. متغیرهای مستقل انتخابی آنان که متغیرهای کلان اقتصادی هستند عبارت اند از شاخص کل قیمت بورس تهران، نرخ ارز در بازار آزاد، قیمت نفت و قیمت طلا. نتایج این پژوهش ضمن نشان دادن توانایی هر دو مدل در پیش‌بینی رفتار بازده سهام، گویای برتری عملکرد شبکه‌های عصبی مصنوعی بر مدل چند عاملی بود.

سینایی، مرتضوی و تیموری اصل (۱۳۸۴) با استفاده از مدل شبکه عصبی مصنوعی MLP و مدل ARIMA شاخص بورس اوراق بهادار تهران را پیش‌بینی نمودند. نتایج این پژوهش نشان داد، شبکه‌های عصبی عملکرد بهتری نسبت به مدل خطی ARIMA برای پیش‌بینی شاخص قیمت دارند. نمازی و کیامهر (۱۳۸۶) بازده روزانه سهام را با استفاده از شبکه‌های مصنوعی پیش‌بینی کردند. در این پژوهش از سه متغیر شاخص سهام، حجم سهام مبادله شده و آخرین نرخ سهام در روز مورد استفاده قرار گرفت.

آن‌ها نشان دادند، رفتار سری زمانی بازده روزانه سهام شرکت‌ها یک فرآیند تصادفی نیست و حافظه دارد. هم‌چنین شبکه‌های عصبی مصنوعی توانایی پیش‌بینی بازده روزانه را با میزان خطای به نسبت مناسبی دارند.

عادل آذر و سیروس کریمی (۱۳۸۸) با استفاده از شبکه‌های عصبی مصنوعی و نسبت‌های حسابداری به پیش‌بینی بازده سهام شرکت‌های فعال در صنعت سیمان و دارو پرداختند و نتایج حاصله را با مدل خطی رگرسیون حداقل مربعات (OLS) مقایسه نمودند. نتایج آنان بیانگر خطای پیش‌بینی کمتر در تکنیک شبکه‌های عصبی مصنوعی بود.

مجتبی مسعودی نژاد و ادهم بیکی خشک با استفاده از شبکه عصبی مصنوعی چندجمله‌ایی (PNN) اقدام به پیش‌بینی بازده سهام شرکت ایران خودرو کردند و مزایا و توانایی‌های این مدل را نشان دادند.

زرانژاد، احمدی‌فرد و ساده‌نژاد (۱۳۹۱) با استفاده از سیستم استنتاج فازی و شبکه‌های عصبی مصنوعی اقدام به پیش‌بینی بازده سهام شرکت ایران خودرو با استفاده از متغیرهای سری زمانی بازده سهام، قیمت جهانی نفت، قیمت جهانی طلا، نرخ ارز و شاخص بورس کردند. پژوهش آنان که مبتنی بر مقایسه دو مدل شبکه عصبی مصنوعی با معماری پس انتشار خطا از نوع پرسپترون چندلایه و شبکه عصبی فازی با الگوریتم یادگیری پس انتشار خطا بود، نشان داد که مدل فازی از دقت بالاتری برخوردار است.

احمد مدرس و محمدباقر کهن‌سال (۱۳۹۲) در پژوهش خود در مورد "ساختارهای خطی و غیرخطی در پیش‌بینی بازده سهام" ساختارهای خطی پیش‌بینی کننده را در قالب دو مدل قیمت‌گذاری دارایی‌های سرمایه‌ایی و سه عاملی فاما و فرنچ و ساختارهای غیرخطی را به صورت شبکه‌های عصبی تشریح می‌نماید. نتایج پژوهش آنان نشان داد که مدل‌های شبکه عصبی مصنوعی به دلیل داشتن ساختاری سفارشی، از انعطاف‌پذیری بالاتری در حل مسائل برخوردار هستند. محدودیت‌های مدل‌های ساختاری موجب شده است تا تحلیل‌گران پیش از مدل‌سازی، نیازمند پیش‌فرض‌هایی برای ساده نمودن روابط واقعی حاکم بر متغیرها باشند در حالی که با شبکه‌های عصبی می‌توان از جایگاهی فراتر از این فرضیات

دست و پاگیر به تحلیل متغیرها و پیش‌بینی رفتار آتی آن‌ها اقدام نمود، بنابراین این روش ابزاری سودمند در پیش‌بینی‌های مالی به حساب آمده و دقت بالاتری به‌دست می‌دهند.

علاوه بر پیش‌بینی بازده سهام، تحقیقات بسیاری در زمینه پیش‌بینی شاخص بورس اوراق بهادار تهران انجام شده است که از آن نمونه می‌توان به موارد زیر اشاره کرد.

محسن مهرآرا و همکاران (۱۳۸۸) با استفاده از ساختار تلفیقی الگوریتم ژنتیک با رویکرد شبکه عصبی GMDH، شاخص بورس اوراق بهادار تهران را پیش‌بینی و عوامل مؤثر بر آن را بیان کردند. نتایج به‌دست آمده نشان دهنده تأثیر قوی و معنادار شاخص قیمت زمین، هزینه مسکن، CPI، پایه پولی، کرایه مسکن اجاره‌ایی و قیمت جهانی نفت خام بر شاخص قیمت و بازده نقدی بورس اوراق بهادار تهران است. در مقابل بازار ارز خارجی و طلا ارتباط کمتری با بازار سهام داشته است.

رضا تهرانی و سعید مرادپور (۱۳۹۱) عملکرد شبکه عصبی شعاع پایه و شبکه عصبی پرسپترون را برای پیش‌بینی بازده شاخص بورس اوراق بهادار با آزمون حداقل مربعات خطا در رو رویکرد درون نمونه‌ایی و برون نمونه‌ایی مورد سنجش قرار دادند. نتایج این پژوهش در رویکرد درون نمونه‌ایی برتری شبکه عصبی شعاع پایه و در رویکرد برون نمونه‌ایی برتری شبکه عصبی پرسپترون را نمایش می‌دهد.

۲-۲-۳-۲ پژوهش‌های انجام گرفته در خارج از کشور

با توجه به گرایش روز افزون علم مدیریت مالی و اقتصاد به روش‌های داده‌کاوی و استفاده از آن‌ها، روش راف ست رونق چشم‌گیری داشته است. از این تکنیک به منظور پیش‌بینی احتمال ورشکستگی یک شرکت، بازاریابی بر مبنای پایگاه داده و سرمایه‌گذاری‌های مالی استفاده می‌شود[۴۱]. در عین حال تحقیقات مالی انجام شده با استفاده از راف ست در خارج از کشور عبارت‌اند از:

تی و شن[1] (۲۰۰۲) طی مقاله خود در زمینه کاربرد راف ست در حیطه اقتصاد و مالی بیان کردند که این تکنیک با توجه به قدرتی که در پردازش جدول اطلاعات حاوی چند ویژگی دارد، ابزار مناسبی برای پیش‌بینی‌های مالی می‌باشد؛ چرا که داده‌های تاریخی در این حوزه در قالب جداول نام برده شده، ظاهر می‌شوند[۴۲]. شن و لوح[2] (۲۰۰۴) با استفاده از روش راف ست روند شاخص S&P 500 را پیش‌بینی کردند و به مشکلات تبدیل سری‌های زمانی، انتخاب شاخص مناسب و ساختن سیستم معاملاتی اشاره نمودند[۴۳]. پس از شن و لوح، تن و کواک[3] در سال ۲۰۰۵ با استفاده از روش RSPOP[4] اقدام به طراحی سیستم معاملاتی کردند[۴۴] و یک سال بعد آنگ و کواک[5] با استفاده از همان روش سیستم معاملاتی را طراحی کردند که اختلاف قیمت میان مدل پیش‌بینی شده و مدل تصمیم‌گیری معاملاتی فاقد تنگراه را مضاعف کند. نتایج این پژوهش بیان‌گر آن بود که سود حاصل از این مدل بیش از سود حاصل از مدل DENFIS[6] و یا سود معامله‌گری بدون در اختیار داشتن مدلی برای پیش‌بینی است[۴۵]. چانگ و همکاران[7] (۲۰۰۶) با تلفیق تئوری راف ست و سیستم خاکستری[8] اقدام به ارائه مدلی کردند که روند سهام بازار بورس تایوان را پیش‌بینی و در انتخاب پرتفلیو بهینه به آنها کمک کند[۴۱]. زاییینی و همکاران[9] (۲۰۰۸) با بکارگیری راف ست تلاش کردند تا عملکرد آتی شرکت‌ها را با استفاده از نسبت‌های نسبت‌های مالی آن‌ها پیش‌بینی کنند، که از این طریق بتوانند بازده سهام آن شرکت‌ها را در آینده تخمین بزنند. آنان شرکتی که بازده سهام آن بیش از بازده شاخص بورس اوراق بهادار کوالالامپور بود را به عنوان شرکتی با عملکرد بالا مدنظر قرار دادند[۴۶]. هوآنگ[10] که محققی در زمینه مالی است، از راف ست به منظور ساخت سیستم اتوماتیک معاملاتی و انتخاب پرتفوی بهینه، بسیار استفاده کرده و به

[1] : Francis E.H. Tay & Lixiang Shen
[2] : Lixiang Shen & Han Tong Loh
[3] : Andy Tan & C. Quek
[4] : rough set-based pseudo outer-product (RSPOP)
[5] : Kai Keng Ang & Chai Quek
[6] : dynamic evolving neural-fuzzy inference system (DENFIS)
[7] : Ting-Cheng Chang, Chuen-Jiuan Jane & Yuan-Paio Lee
[8] : Grey System Theory
[9] : Bahtiar Jamili Zaini, Siti Mariyam Shamsuddin & Saiful Hafizah Jaaman
[10] : Kuang Yu Huang

منظور ارتقای سطح عملکرد این سیستم دست به تلفیق تئوری راف ست با بسیاری از روش‌ها زده است. پژوهش‌های او در این زمینه بدین شرح است: هوآنگ و همکاران[1] (۲۰۰۸) از مدل وزن‌دهی تجمعی[2] که که شامل الگوریتم تجمعی میانگین K[3] و مدل خاکستری[4] بود به منظور پیش‌بینی روند آتی و طبقه‌بندی طبقه‌بندی سهام بورس اوراق بهادار تایوان استفاده کردند. روش راف ست را برای انتخاب سهام مناسب و سیستم خاکستری را برای وزن‌دهی بکار بستند[۴۷]. هوآنگ و همکاران[5] (۲۰۰۸) مدل میانگین هندسی[6] هندسی[6] جدیدی به منظور افزایش کارایی سیستم تجزیه و تحلیل ارتباط خاکستری[7] بکار گرفتند. روش روش میانگین هندسی تجزیه و تحلیل ارتباط خاکستری[8]، تلفیقی از مدل پیش‌بینی خود توضیح غیرخطی با داده‌های بیرونی[9]، تئوری سیستم خاکستری و راف ست است که بتواند مکانیزم اتوماتیکی به منظور پیش‌بینی بازار بورس اوراق بهادار ارائه دهد. نتایج حاصل از پرتفلیوی تشکیل شده با این روش را با روش تخصیص صندوق میانگین وزنی[10] مقایسه کردند و نتایج حاکی از آن بود که روش جدید نرخ بازده بالاتری را منجر می‌شود[۴۸]. هوآنگ و جین[11] (۲۰۰۸) برای ساخت مکانیزم اتوماتیک پیش‌بینی روند سهام و تشکیل پرتفوی بهینه از درجه ارتباط خاکستری[12] استفاده کردند. این مدل تلفیقی است از روش پیش‌بینی ARX، تئوری GM و مدل راف ست. نتایج حاکی از آن بود که این سیستم اتوماتیک نرخ بازده متوسط سالانه ۲۵.۹۱ درصد را برای سهام انتخاب شده از بورس اوراق بهادار تایوان منجر می‌شود[۴۹]. هوآنگ و جبن (۲۰۰۸) به منظور ساخت مکانیزم اتوماتیک انتخاب پرتفلیو از تئوری طبقه‌بندی فازی میانگین C[13] و هم‌چنین تئوری‌های فازی به منظور انتخاب مقدار مناسب برای متغیر

[1] : Kuang Yu Huang, Chuen-Jiuan, Jane & Ting-Cheng Chang
[2] : Weight Clustering Model
[3] : K-means Clustering Algorithm
[4] : Grey Model: GM(1,N)
[5] : Kuang Yu Huang, Shu-Ling Shieh, Chuen-Jiuan Jane & De Jyun Jheng
[6] : Geometry Mean (GM) Model
[7] : Gray Relational Analysis (GRA)
[8] : GM method of GRA
[9] : AutoRegressive with exogenous inputs (ARX)
[10] : average weight fund allocation method
[11] : Kuang Yu Huang & Chuen-jiuan Jane
[12] : Grey Relational Grade (GRG)
[13] : Fuzzy C-Means (FCM) classification theory

آستانه‌ایی راف ست دقیق (β)[1] در موارد مواجهه با طبقه‌بندی سیستم اطلاعات پیوسته، استفاده کردند، سپس مدل راف ست با دقت متغیر(VPRS)[2] را با مدل پیش‌بینی میانگین متحرک خود توضیح بیرونی (ARX)[3] و سیستم خاکستری ادغام کردند. در این روش داده‌های مالی به ARX وارد می‌شوند تا روند آتی آنها پیش‌بینی شود. سپس توسط مدل میانگین هندسی تقلیل می‌یابند و به وسیله الگوریتم طبقه‌بندی فازی میانگین C دسته‌بندی می‌شوند. به منظور انتخاب سهام مناسب برای سرمایه‌گذاری، با توجه به قوانینی که تصمیم‌گیرندگان از پیش تعیین کرده‌اند، از VPRS استفاده شده است و در نهایت به منظور به حداکثر رسانیدن بازده پرتفلیو سیستم تجزیه و تحلیل ارتباط خاکستری[4] بکار گرفته شده است است تا سهام را وزن‌دهی کند. نتایج حاصل از این تحقیق نشان می‌دهد که متوسط نرخ بازده سالانه پرتفلیو تشکیل شده به وسیله این روش برای سهام منتخب اوراق بهادار بورس تایوان بین سال‌های۲۰۰۴ تا ۲۰۰۶، ۲۹.۳۳ درصد است[۵۰]. یک سال بعد هوآنگ نتایج حاصل از مقاله قبل را با روش مدل انتخابی راف ست مقایسه کرد و به این نتیجه رسید که بازده مدل ترکیبی به مراتب بیشتر از مدل انتخابی راف ست بوده است[۵۱]. هوآنگ و جین (۲۰۰۹) با تلفیق سه روش راف ست، سیستم‌های خاکستری و مدل خود توضیح میانگین متحرک بیرونی اقدام به ارائه مدلی کردند که روند سهام بازار بورس تایوان را پیش‌بینی و در انتخاب پرتفلیو مناسب به آنها کمک کند. نتایج حاصل از این مکانیزم را با نتایج حاصل از مدل پیش‌بینی خاکستری مقایسه کردند و بیان کردند که روش ترکیبی نه تنها بازده بیشتری را منجر می‌شود، بلکه از دقت بالاتری نیز برخوردار است[۵۲]. هوآنگ و وان[5] (۲۰۱۱) با استفاده استفاده از راف ست، نمرات رابطه خاکستری[6] و روش شاخص PBMF، ۵۳ شاخص مالی مربوط به هر سهم را با بکارگیری تجزیه و تحلیل روابط خاکستری در شش نسبت مالی، یکی کردند. روش تجزیه و تحلیل ارتباط خاکستری وزن بهینه سهام انتخاب شده را به نحوی تعیین میکند که نرخ بازده پرتفلیو به

[1] : Variable Precision Rough Set (VPRS) threshold parameter (β)
[2] : Variable Precision Rough Set (VPRS)
[3] : Moving Average Autoregressive Exogenous (ARX)
[4] : Grey relational analysis technique
[5] : Kuang Yu Huang & Shiuan Wan
[6] : Grey relational grades (GRGs)

حداکثر برسد. نتایج حاصل از این پژوهش نشان می‌دهد که بازده این تکنیک از بازده بسیاری از روش‌ها بالاتر است[۵۳].

یآاو و هربرت[1] (۲۰۰۹) طی مقاله خود در مورد تجزیه و تحلیل سری‌های زمانی مالی با روش راف ست بدین نتیجه رسیدند که این روش تکنیک مناسبی برای تحلیل داده‌های مبهم و ناقص است و روند بورس اوراق بهادار نیوزیلند را با استفاده از آن پیش‌بینی کردند و معتقدند نتایج دقیق و قابل اتکایی کسب کرده‌اند[۵۴].

جامان و همکاران[2] (۲۰۰۹) با استفاده از تئوری راف ست روند سهام بورس اوراق بهادار مالزی را به منظور شناخت زمان مناسب برای خرید و فروش اوراق بهادار پیش‌بینی کردند[۵۵] و در همان سال با استفاده از ترکیب راف ست با شبکه عصبی مصنوعی نیز این کار را تکرار نمودند و نشان دادند که این روش در ساده‌سازی سری‌های زمانی و مدیریت اطلاعات ناقص توانمند می‌باشد[۵۶]. القاهری و همکاران[3] (۲۰۰۹) طرحی کلی برای استخراج قوانین و اصول پیش‌بینی، با استفاده از راف ست در بورس اوراق بهادار کویت طراحی کردند. آن‌ها نتایج حاصل از راف ست را با نتایج شبکه عصبی مصنوعی و درخت تصمیم‌گیری مقایسه کردند و دریافتند که دقت راف ست در پیش‌بینی، به ۹۷ درصد می‌رسد و تعداد قوانین کمتری نسبت به دو روش نام برده شده، بکار می‌گیرد[۵۷]. چاودهوری و همکاران[4] (۲۰۰۹) با استفاده از روش شبکه‌های مفهومی چند لایه راف ست فازی[5] قیمت سهام بورس اوراق بهادار بمبی را پیش‌بینی کردند. آنها از راف ست به منظور ساخت شبکه مناسبی هم برای طبقه‌بندی و هم برای استخراج قوانین استفاده کردند. نتایج حاکی از آن بود که این روش با کمترین تعداد قوانین استخراج شده، نتایج را با دقتی بالا و زمان محاسبه پایین و به صورت واضح و روشن ارائه می‌دهد[۵۸]. چنگ و

[1] : JingTao Yao & Joseph P. Herbert
[2] : Saiful Hafizah Jaaman et al.
[3] : Hameed Al-Qaheri, Aboul Ella Hassanien & Ajith Abraham
[4] : Arindam Chaudhuri, Kajal De & Dipak Chatterjee
[5] : Rough Fuzzy Multi Layer Perception Networks

همکاران[1] (۲۰۱۰) با تلفیق روش راف ست و شبکه عصبی مصنوعی اقدام به پیش‌بینی قیمت سهام بورس تایوان و همچنین ارائه مدلی برای زمان‌بندی سرمایه‌گذاری کردند. برای نیل به این هدف سه مدل طراحی کردند. مدل اول با استفاده از شبکه عصبی، مدل دوم راف ست و مدل سوم به صورت تلفیقی. نتایج حاکی از آن بود که مدل تلفیقی برآورد بهتر با درصد خطای کمتری ارائه می‌دهد[۵۹]. چنگ و همکاران[2] در همان سال با استفاده از روش ترکیبی راف ست و الگوریتم ژنتیک به پیش‌بینی قیمت سهام بورس تایوان با استفاده از تجزیه و تحلیل تکنیکال پرداختند[۶۰]. بانیک و همکاران[3] (۲۰۱۴) با استفاده از روش راف ست و شبکه عصبی مصنوعی و تلفیقی از هر دو، مدلی برای پیش‌بینی بازار بورس بنگلادش طراحی کردند و بدین نتیجه رسیدند که مدل تلفیقی بهترین روش برای نیل به این هدف است[۶۱]. سعید حسن[4] (۲۰۱۴) با استفاده از راف ست، روند بازار مصر را پیش‌بینی کرد. به عقیده او پیش‌بینی روند بازار بورس و تعیین زمان خرید و فروش سهام به مراتب مؤثرتر از پیش‌بینی قیمت آنها است. در این پژوهش سرمایه‌گذاران به پنج دسته طبقه‌بندی شده‌اند: مصریان، عرب‌ها، افراد خارجی، افراد معمولی و مؤسسات. هر دسته از سرمایه‌گذاران رفتار منحصر به فردی از خود نشان می‌دهند و مطمئناً معاملات آنها به حالت‌های مختلفی از قیمت دارایی مربوط می‌باشد و نوع ارتباط این گروه‌ها با تغییرات قیمت بازار که با استفاده از تغییرات در روند آن حاصل می‌شود، از لحاظ هم‌زمانی یا متاخر بودن بررسی شده است. نتایج مهر تاییدی بر قدرت راف ست در زمینه استخراج حقایق از دل اطلاعات پنهان و همچنین توانایی آن در طبقه‌بندی داده‌ها زد[۶۲].

[1] : Jao-Hong Cheng, Huei-Ping Chen & Yi-Min Lin
[2] : Ching-Hsue Cheng, Tai-Liang Chen & Liang-Ying Wei
[3] : Shipra Banik, A F M Khodadad Khan & Mohammad Anwer
[4] : Naglaa Ragaa Saeid Hassan

فصل سوم

متدولوژی تحقیق

۳-۱ مروری اجمالی بر داده‌کاوی

۳-۱-۱ مقدمه :

در طول سالیان اخیر، حجم مخازن داده‌های الکترونیکی ایجاد شده توسط بانک‌ها، شرکت‌ها و سایر مؤسسات تجاری، همواره در حال افزایش بوده است. بیت‌های با ارزش اطلاعات در این مخازن داده، جای گرفته‌اند، به گونه‌ای که اندازه و حجم بسیار زیاد این منابع، تجزیه و تحلیل آن‌ها را برای انسان به منظور خلق اطلاعات یا الگوهای مفید و مناسب در جهت فرآیند تصمیم‌گیری، غیرممکن ساخته است. فنونی نظیر داده‌کاوی می‌توانند به طور خودکار، اطلاعات را از میان این مقادیر حجیم داده استخراج نمایند. داده‌کاوی تأثیراتی عمیق بر شیوه‌های کسب و کار و مدیریت دانش در سال‌های اخیر داشته است. هوشمندی کسب و کار، مشهورترین کاربرد فنون داده‌کاوی است. قابلیت‌های طبقه‌بندی و پیش‌بینی ابزارهای داده‌کاوی، موجب کاربرد این ابزارها به منظور مقاصد پیش‌بینی ورشکستگی، وضعیت تداوم فعالیت، پیش‌بینی درماندگی مالی، تشخیص و شناسایی تقلب مدیریت، هم‌چنین تخمین ریسک اعتباری و عملکرد واحد تجاری شده و بدین ترتیب داده‌کاوی را تبدیل به موضوعی با درجه اهمیت زیاد، در امور مالی و حسابداری نموده است.

در دنیای تجارت نیز داده‌های مالی به عنوان سرمایه راهبرد مطرح هستند. داده‌های مالی توسط مؤسساتی مثل بانک‌ها، بورس اوراق بهادار، سازمان‌های مالیاتی، پایگاه‌های داده ویژه حسابرسان و حسابداران و غیره جمع‌آوری و نگهداری می‌شوند. روش‌های داده‌کاوی در داده‌های مالی، می‌تواند در حل مشکلات طبقه‌بندی و تسهیل فرآیند تصمیم‌گیری به کار روند. نمونه‌هایی از مسائل طبقه‌بندی مالی[۱] شامل ورشکستگی شرکت‌ها[۲]، تخمین ریسک اعتباری[۳]، گزارش تداوم فعالیت[۴]، درماندگی مالی[۵]

---

1 : Financial Classification Problem
2 : Corporate Bankruptcy
3 : Credit Risk Estimation
4 : Going Concern Reporting
5 : Financial Distress

و پیش‌بینی عملکرد واحد تجاری[1] می‌باشد. بدین ترتیب دامنه اهمیت داده‌کاوی در امور مالی و حسابداری می‌تواند طیفی گسترده از کشف تقلب تا افزایش سودآوری واحد تجاری باشد.

اهمیت داده‌کاوی توسط بسیاری از سازمان‌های حرفه‌ای تشخیص داده شده است. انجمن حسابداران رسمی آمریکا، داده‌کاوی را به عنوان یکی از ده فناوری برتر برای آینده معرفی کرده است. هم‌چنین انجمن حسابرسان داخلی آمریکا نیز این فناوری را در فهرست یکی از چهار اولویت تحقیقاتی خود گنجانده است [۶۳].

۳-۱-۲ پیشینه داده‌کاوی :

داده‌ها نمایشی از واقعیت‌ها، معلومات، مفاهیم، رویدادها یا پدیده‌ها برای برقراری ارتباط، تفسیر یا پردازش، توسط انسان یا ماشین می‌باشند. از طرف دیگر واژه‌ی اطلاعات به دانشی که از طریق خواندن، مشاهده و آموزش به دست می‌آید اطلاق می‌شود و در حقیقت می‌توان گفت اطلاعات داده‌هایی هستند که پس از جمع‌آوری پردازش شده‌اند و شکل مفهومی آن را دانش می‌نامند. بین داده‌ها و اطلاعات یک شکاف وجود دارد که اندازه‌ی این شکاف با حجم داده‌ها ارتباط مستقیم دارد. هرچه داده‌ها حجیم‌تر باشند، این شکاف بیشتر خواهد بود و هر چه حجم داده‌ها کمتر و روش‌ها و ابزارهای پردازش داده‌ها کاراتر باشد، فاصله‌ی بین داده‌ها و اطلاعات کمتر است. امروزه افزایش سریع حجم پایگاه داده به شکلی است که توانایی انسان برای درک این داده‌ها بدون ابزارهای پرقدرت میسر نمی‌باشد. در این وضعیت، تصمیم‌گیری‌ها به جای تکیه بر اطلاعات بر درک مدیران و کاربران تکیه دارند، چرا که تصمیم گیرندگان ابزار قوی برای استخراج اطلاعات با ارزش را در دست ندارند.

از سال ۱۹۵۰ رایانه‌ها در تحلیل و ذخیره‌سازی داده‌ها به کار گرفته شدند. پس از حدود ۲۰ سال حجم داده‌ها دو برابر شد و پس از آن تقریباً هر دو سال یک بار هم‌زمان با پیشرفت فناوری اطلاعات، حجم داده‌ها هم به دو برابر افزایش می‌یافت. این پیشرفت آن قدر زیاد بود که تعداد رکوردهای برخی

[1] : Corporate Performance Prediction

از پایگاه داده‌ها به چند صد میلیارد رسید و نیاز به تحلیل این حجم عظیم از داده افزایش یافت.

پیشرفت شگفت‌انگیز فناوری رایانه‌ای و مجهز شدن بشر به این ابزار سبب پیشرفت فوق‌العاده در کسب و ذخیره‌سازی داده‌های عددی و هم‌چنین به وجود آمدن پایگاه داده‌های بزرگ در زمینه‌های مختلف شده است. داده‌های تبادلات تجاری، کشاورزی، ترافیک، اینترنت، داده‌های نجومی، جزئیات مکالمات تلفنی، داده‌های پزشکی و درمانگاهی مثال‌هایی از چنین پایگاه داده‌هایی می‌باشند. در واقع تکنیک‌های تولید و جمع‌آوری پایگاه داده بسیار سریع‌تر از توانایی بشر در درک و استفاده از آن‌ها رشد کرده است. از اواخر دهه‌ی ۸۰ میلادی بشر به فکر دستیابی به اطلاعات نهفته در این داده‌های حجیم افتاد و تلاش‌ها برای انجام این کار را شروع کرد که با سیستم‌های سنتی استفاده از پایگاه داده میسر نبود. شدت رقابت‌ها در عرصه‌های علمی، اجتماعی، اقتصادی، سیاسی و نظامی نیز اهمیت عامل سرعت یا زمان دسترسی به اطلاعات را دو چندان کرد. بنابراین نیاز به طراحی سیستم‌هایی که قادر به اکتشاف سریع اطلاعات مورد علاقه‌ی کاربران با تأکید بر حداقل مداخله‌ی انسانی باشند از یک طرف و روی آوردن به روش‌های تحلیل متناسب با حجم داده‌های زیاد از سوی دیگر احساس شد.

در سال ۱۹۸۹ و ۱۹۹۱ کارگاه‌های کشف دانش و معرفت از پایگاه‌های داده توسط پیاتتسکی[۱] و همکاران برگزار شد. در واقع داده‌کاوی فرآیندی است که در آغاز دهه‌ی ۹۰ پا به عرصه‌ی ظهور گذاشته و با نگرشی نو، به مسأله‌ی استخراج اطلاعات از پایگاه داده‌ها پرداخت. در واقع پژوهش جدی روی موضوع داده‌کاوی از اوایل دهه‌ی ۹۰ شروع شد. پژوهش‌ها و مطالعه‌های زیادی در این زمینه صورت گرفت؛ همچنین سمینارها، دوره‌های آموزشی و کنفرانس‌هایی نیز برگزار شد و پایه‌های نظری داده‌کاوی در تعدادی از مقاله‌های پژوهشی آورده شد. در فواصل سال‌های ۱۹۹۱ تا ۱۹۹۴ کارگاه‌های کشف دانش و معرفت از پایگاه داده‌ها توسط فیاد[۲] و پیاتتسکی و دیگران برگزار شد. از سال ۱۹۹۵ داده‌کاوی به صورت جدی وارد مباحث آمار شد.

واژه‌ی «کشف دانش» به طور رسمی اولین بار توسط فیاد در اولین کنفرانس بین‌المللی داده‌کاوی و

---

۱ :G.Piatetsky-Shapiro

۲ :Usama Fayyad

کشف دانش که در سال ۱۹۹۵ در مونترال کانادا برگزار شده بود، معرفی شد که به بیان ارتباط تکنیک‌های آنالیز در چندین مرحله با هدف استخراج دانش‌های ناشناخته‌ی قبلی از داده‌های در دسترس می‌پرداخت. داده‌هایی که ارتباط منظم و پراهمیت آن‌ها قبلاً به نظر نمی‌رسید.

گروهی از محققین سال ۱۹۹۵ با استفاده از داده‌کاوی، انبارهای داده‌ی بانک‌های آمریکا را بررسی کرده و بیان کردند که چگونه این سیستم ها برای بانک‌های آمریکا قدرت رقابت بیشتری ایجاد می‌کنند. در این سال انجمن داده‌کاوی هم‌زمان با اولین کنفرانس بین‌المللی «کشف دانش و داده‌کاوی» شروع به کار و یک سازمان علمی به نام ACM- SIGKDD را تاسیس کرد. در سال ۱۹۹۶ اولین شماره‌ی مجله‌ی «کشف دانش از پایگاه داده‌ها» منتشر شد. در همان سال دیدگاهی از داده‌کاوی به عنوان «پرس و جو کننده از پایگاه‌های استنتاجی» پیشنهاد شد و فیاد و پیاتتسکی پیشرفت‌های کشف دانش و داده‌کاوی را اعلام کردند. همچنین دیدگاه اقتصادسنجی روی داده‌کاوی و عملکرد داده‌کاوی به عنوان یک مسأله‌ی بهینه، ارائه و کنفرانس‌های ناحیه‌ای و بین‌المللی در مورد داده‌کاوی برگزار شد که از آن جمله می‌توان به کنفرانس آسیا و اقیانوسیه درباره‌ی کشف دانش و داده‌کاوی اشاره کرد. در سال ۲۰۰۰ بحث‌های مقایسه‌ای بین آمار و داده‌کاوی و نیز استفاده از وب در کاوش داده‌ها و کاربردهای آن ارائه شد و در نهایت در سال ۲۰۰۲ «داده‌کاوی ساختارهای پیوند برای مدل رفتار مصرف کننده» عرضه شد. در حال حاضر، داده‌کاوی مهمترین فناوری جهت بهره‌برداری مؤثر از داده‌های حجیم است و اهمیت آن رو به افزایش است [۶۶,۶۵,۶۴].

۳-۱-۳ مفاهیم کلی داده‌کاوی

تکامل تکنولوژی پایگاه داده و استفاده فراوان آن در کاربردهای مختلف سبب جمع‌آوری حجم فراوانی داده شده است. این داده‌های فراوان باعث ایجاد نیاز برای ابزارهای قدرتمند به منظور تحلیل داده‌ها گشته، زیرا در حال حاضر به لحاظ داده ثروتمند هستیم ولی دچار کمبود اطلاعات می‌باشیم.

ابزارهای داده‌کاوی داده‌ها را آنالیز و الگوهای آنان را کشف می‌کنند. شکاف موجود بین داده‌ها و

اطلاعات سبب ایجاد نیاز برای بازارهای داده‌کاوی شده است تا داده‌های بی‌ارزش را به دانشی ارزشمند تبدیل کنیم [۶۷].

داده‌کاوی عبارت است از فرآیند یافتن دانش از مقادیر عظیم داده‌های ذخیره شده در پایگاه داده انبار داده و یا دیگر مخازن اطلاعات [۶۸]. در یک تعریف غیررسمی داده‌کاوی فرآیندی است خودکار برای استخراج الگوهایی که دانش را بازنمایی می‌کنند، که این دانش به صورت ضمنی در پایگاه داده‌های عظیم، انبار داده و دیگر مخازن بزرگ اطلاعات، ذخیره شده است.

واژه داده‌کاوی و کشف دانش در پایگاه داده[1]، اغلب به صورت مترادف یکدیگر مورد استفاده قرار می‌گیرند، در حالی که داده‌کاوی بخشی از فرآیند کشف دانش محسوب می‌شود. همان‌طور که در شکل شماره (۳-۱) مشاهده می‌شود، کشف دانش در پایگاه داده فرآیند شناسایی درست، ساده و مفید الگوها و مدل‌های قابل فهم در داده‌ها می‌باشد ولی داده‌کاوی مرحله‌ای از فرآیند کشف دانش است و شامل الگوریتم‌های مخصوص داده‌کاوی است، بطوری که تحت محدودیت‌های مؤثر محاسباتی قابل قبول، الگوها و یا مدل‌ها را در داده کشف می‌کند. به بیان ساده‌تر، داده‌کاوی به فرآیند استخراج دانش ناشناخته، درست و بالقوه از داده اطلاق می‌شود. داده‌کاوی گام پنجم از مراحل هفت‌گانه کشف دانش را تشکیل می‌دهد:

۱-پاکسازی داده‌ها[2] (از بین بردن نویز و ناسازگاری داده‌ها).

۲-یکپارچه‌سازی داده‌ها[3] (چندین منبع داده ترکیب می‌شوند).

۳-انتخاب داده‌ها[4] (داده‌های مرتبط با آنالیز از پایگاه داده بازیابی می‌شوند).

۴-تبدیل کردن داده‌ها[5] (تبدیل داده‌ها به فرمی که مناسب برای داده‌کاوی باشد مثل خلاصه‌سازی و همسان‌سازی[6]).

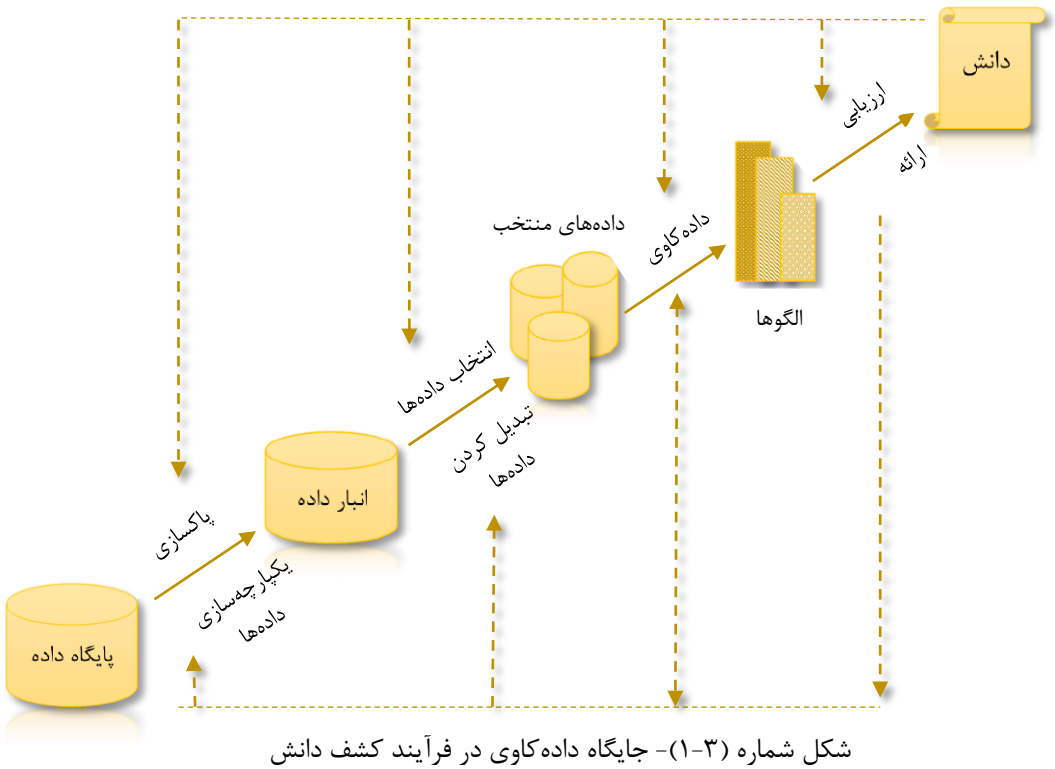
---

[1] : KDD: Knowledge Discovery of Database
[2] : Data Cleaning
[3] : Data Integration
[4] : Data Selection
[5] : Data Transformation
[6] : Summary & Aggregation

۵-داده‌کاوی (فرآیند اصلی که روال‌های هوشمند برای استخراج الگوها از داده‌ها به کار گرفته می‌شوند).

۶-ارزیابی الگو[1] (برای مشخص کردن الگوهای صحیح و مورد نظر به وسیله معیارهای اندازه‌گیری).

۷-ارائه دانش[2] (یعنی نمایش بصری، تکنیک‌های بازنمایی دانش برای ارائه دانش کشف شده به کاربر استفاده می‌شود).



شکل شماره (۳-۱)- جایگاه داده‌کاوی در فرآیند کشف دانش

فرآیند داده‌کاوی خود به تنهایی شامل چندین مرحله می‌باشد که به صورت جزئی عبارت‌اند از:

۱- شناسایی هدف: در این مرحله مشخص می‌شود کاربر به دنبال کشف چه موضوعی است و قصد دارد تا چه سطحی از اطلاعات را در اختیار داشته باشد که با توجه به آن، اطلاعات مورد نیاز را از پایگاه داده اخذ نماید.

۲- انتخاب داده‌ها: در این مرحله با توجه به معیارهای مشخص شده، داده‌ها انتخاب می‌گردند.

[1] : Pattern Evaluation
[2] : Knowledge Presentation

۳- آماده‌سازی داده‌ها: شکل قابل استفاده داده و شناسایی متغیرهای زائد وظیفه این مرحله از فرآیند خواهد بود.

۴- ارزیابی داده‌ها: چارچوب کلی این مرحله، معیارهایی از قبیل نوع توزیع داده‌ها، ویژگی‌ها و ساختار پایگاه داده و شرایط کلی داده‌ها و .... می‌باشد.

۵- قالب‌بندی پاسخ: خروجی ان بخش، اراده اطلاعات در قالب‌های مختلف و قابل فهم ساده همچون نمودارها، تصاویر، شبکه‌ها عصبی و .... است.

۶- انتخاب ابزار: در این مرحله ابزار مناسب برای داده‌کاوی انتخاب می‌گردد.

۷- الگوسازی: فرآیند داده‌کاوی به صورت اصلی از این مرحله آغاز می‌گردد که شامل جست‌وجوی الگوها در مجموعه داده، طبقه‌بندی و ارزشیابی داده‌ها و ... می‌باشد.

۸- اعتبارسازی یافته‌ها: الگوهای حاصله ارزیابی می‌گردند و میزان اعتبار آن‌ها سنجیده می‌شود.

۹- ارائه نتایج: در این بخش نتایج نهایی فرآیند به کاربر ارائه می‌گردد.

۱۰- استفاده از نتایج: هدف اصلی داده‌کاوی استفاده از نتایج کشف شده برای تصمیم‌گیری، سیاست‌گذاری و پیش‌بینی به منظور ایجاد یک موقعیت بهتر و جدید می‌باشد [۶۹].

فرآیند کلی داده‌کاوی به صورت جزئی در شکل (۳-۲) نشان داده شده است:

شکل شماره (۳-۲)- مراحل فرآیند داده‌کاوی

ریشه‌های داده‌کاوی در میان سه خانواده از علوم، قابل پیگیری می‌باشد. مهمترین این خانواده‌ها آمار کلاسیک[1] است. بدون آمار داده‌کاوی وجود نخواهد داشت، به طوری که این رشته، اساس اغلب تکنولوژی‌هایی می‌باشد که داده‌کاوی بر روی آن‌ها بنا می‌شود. آمار کلاسیک مفاهیمی مانند تحلیل رگرسیون، توزیع استاندارد، انحراف معیار، واریانس، تحلیل خوشه‌ای و فاصله‌های اطمینان را در بر می‌گیرد که همه‌ی این موارد برای مطالعه‌ی داده و ارتباط بین داده‌ها مورد استفاده قرار می‌گیرند. داده‌کاوی جانشین تکنیک‌های آماری سابق نمی‌باشد بلکه وارث آن‌ها بوده و در واقع تغییر و گسترش تکنیک‌های سابق برای متناسب‌سازی آن‌ها با حجم داده‌ها و مسائل امروزی می‌باشد. تکنیک‌های کلاسیک برای داده‌های محدود و مسائل ساده مناسب بوده‌اند حال آن‌که با پیچیده شدن مسائل و رشد روزافزون داده‌ها نیاز به تغییر آن‌ها کاملاً طبیعی است، در هر حال بدون شک تحلیل آماری کلاسیک نقش اساسی در تکنیک‌های داده‌کاوی ایفا می‌کند[۷۰].

دومین خانواده‌ایی که داده‌کاوی به آن تعلق دارد هوش مصنوعی[2] است. هوش مصنوعی که بر پایه روش‌های ابتکاری بنا شده است و با آمار ضدیت دارد، در تلاش است تا فرآیندی مانند ذهن انسان را برای حل مسائل آماری بکار بندد. این رویکرد به دلیل نیاز به توان محاسباتی بالا، تا اوایل دهه ۱۹۸۰ عملی نشده و کاربردهای کمی در حوزه‌های علمی و حکومتی داشته زیرا نیاز به استفاده از کامپیوترهای بزرگ باعث شد همه افراد نتوانند از تکنیک‌های ارائه شده استفاده کنند[۷۱].

سومین خانواده داده‌کاوی، یادگیری ماشین[3] می‌باشد که به مفهوم دقیق‌تر اجتماع آمار و هوش مصنوعی است در حالی که هوش مصنوعی نتوانست موفقیت تجاری کسب کند، یادگیری ماشین در بسیاری از موارد جایگزین آن گردید. از یادگیری ماشین به عنوان تحول هوش مصنوعی یاد شده، زیرا ترکیبی از روش‌های ابتکاری هوش مصنوعی و تحلیل آماری پیشرفته می‌باشد. یادگیری ماشین اجازه می‌دهد تا برنامه‌های کامپیوتری در مورد داده‌ای که آن‌ها مطالعه می‌کنند، مانند برنامه‌هایی که

---

[1] : Classic Statistics
[2] : Artificial Intelligence
[3] : Machine Learning

تصمیم‌های متفاوتی بر مبنای کیفیت داده مطالعه شده می‌گیرند، یادگیری داشته باشند و برای مفاهیم پایه‌ای آن از آمار استفاده می‌کنند و از الگوریتم‌ها و روش‌های ابتکاری هوش مصنوعی برای رسیدن به هدف بهره می‌گیرند. داده‌کاوی در بسیاری از جهات، سازگاری تکنیک‌های یادگیری ماشین با کاربردهای تجاری است. بهترین توصیف از داده‌کاوی بوسیله اجتماع آمار، هوش مصنوعی و یادگیری ماشین بدست می‌آید [۷۲].

مهم‌ترین نکته این است که داده‌کاوی راهکاری است برای مسائل مالی و تجاری امروز به کمک تکنیک‌های آماری و هوش مصنوعی برای افراد حرفه‌ایی که قصد دارند یک مدل پیش‌بینی ایجاد نمایند.

۳-۱-۴ ابزارهای داده‌کاوی

ابزارهای داده‌کاوی نرم‌افزارهایی هستند که به کاربران اجازه استخراج اطلاعات از داده‌ها را می‌دهند. این ابزارها توانایی گردآوری داده‌ها و کاربرد آن‌ها به منظور تصمیم‌گیری در خصوص مصرف‌کننده خاص یا گروهی از مصرف‌کنندگان را به سازمان‌ها و افراد می‌دهند. استخراج دستی داده‌ها از صدها سال پیش وجود داشته است. این در حالی است که ماشینی شدن فرآیند داده‌کاوی از هنگام ورود کامپیوتر شایع شده است. هدف نهایی این ابزارها ظاهر ساختن الگوهای پنهان[1] می‌باشد [۷۳]. به هر حال ابزارهای داده‌کاوی می‌تواند شامل نرم‌افزارهای زیر باشد:

1. Data to Knowledge
2. SAS
3. Clementine
4. Intelligent-miner
5. Insightful miner
6. Weka
7. Microsoft Excel
8. ACL

[1] : Hidden Pattern

۳-۱-۵ انواع روش‌های داده‌کاوی

اصطلاح روش‌های داده‌کاوی در واقع بیان‌گر جمع کثیری از الگوریتم‌ها و فنون است که از علومی مانند آمار، یادگیری ماشین، پایگاه داده و تجسم‌سازی[1] استنتاج شده‌اند. گرچه روش‌های مختلف داده‌کاوی اهداف متفاوتی دنبال می‌کنند، اما در حالت کلی به دو دسته تقسیم می‌شوند : (۱) بازبینی و رسیدگی، (۲) اکتشاف . بازبینی و رسیدگی به اثبات فرضیه‌های کاربران محدود می‌شود در حالی که اکتشاف به دنبال مدل‌های جدید می‌گردد و به دو دسته روش‌های پیش‌بینی‌کننده و توصیف‌کننده تقسیم می‌شود. بسیاری از روش‌های داده‌کاوی در هر دو مجموعه اهداف پیش‌بینی و توصیف‌کنندگی مورد استفاده قرار می‌گیرند. مدل‌های پیش‌بینی از پایگاه داده و بنیان سیستم استخراج می‌شوند و به هدایت رفتار آتی کمک می‌کنند، در حالی که روش‌های توصیف‌کننده این امکان را فراهم می‌کنند که فرآیند تولید داده برای کاربران قابل درک شود. هدف از پیش‌بینی تمرکز بر روی میزان دقت پیش‌بینی و هدف از توصیف تمرکز بر درک فرآیند تولید داده‌هاست. در نمودار شماره (۳-۳) علاوه بر طبقه‌بندی جزئی روش‌های داده‌کاوی، پنج روش تحلیلی آن نیز قابل مشاهده می‌باشد. این پنج روش عبارت‌اند از:

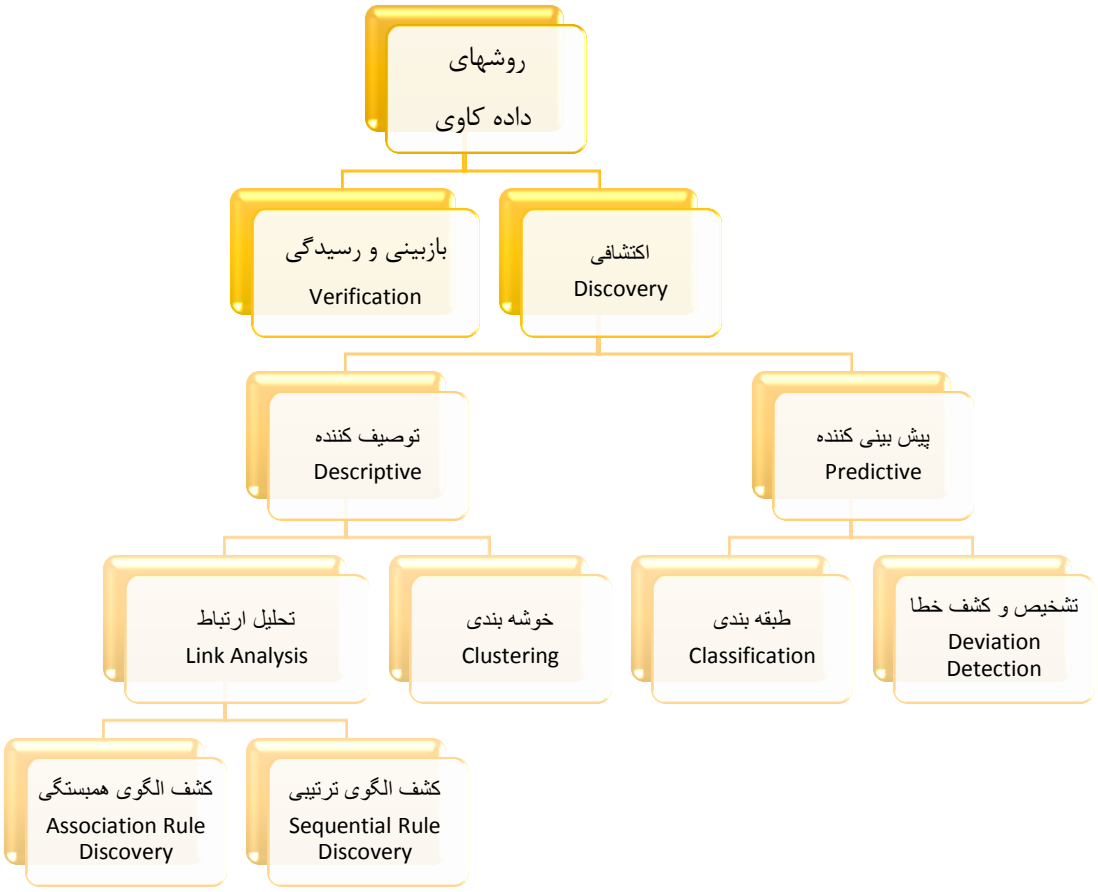
۱- پیش‌بینی خودکار روندها و رفتارها:

داده‌کاوی به صورت اتوماتیک در پایگاه داده به دنبال اطلاعات قابل پیش‌بینی می‌گردد. پیش‌بینی بازار یک مثال ساده در این زمینه می‌باشد که در آن داده‌کاوی بر مبنای داده‌های گذشته به منظور کشف مشتری بالقوه با بالاترین ارزش بازده سرمایه‌گذاری صورت می‌گیرد. سایر موارد پیش‌بینی روند، شامل پیش‌بینی ورشکستگی شرکت‌ها در آینده و انتخاب گروه‌های مربوطه است که احتمال دارد به رویدادهای مشخص پاسخگو باشند.

[1] : Visualization

۲- تجزیه و تحلیل ارتباط

ارتباط و وابستگی داده از مهمترین بخش‌های دانش قابل اکتشاف در پایگاه داده محسوب می‌شود. اگر بین دو یا چند متغیر نظم خاصی برقرار باشد، می‌توان گفت نوعی وابستگی و ارتباط میان آن‌ها وجود دارد. این ارتباط به دو دسته تقسیم می‌شود: (۱) ارتباط سری زمانی، (۲) ارتباط علی. هدف از تجزیه و تحلیل ارتباط، یافتن شبکه‌های مرتبط پنهان در پایگاه داده است. در برخی موارد عدم آگاهی مربوط به عملکرد ارتباطی بین داده‌ها می‌شود و در شرایطی که این آگاهی وجود داشته باشد، نامعلوم است. بنابراین قوانین ایجاد شده به وسیله تجزیه و تحلیل ارتباط، برای نشان دادن احتمال وقوع آنها استفاده می‌شود.



نمودار شماره (۳-۳)- طبقه‌بندی انواع روش‌های داده‌کاوی

۳- تجزیه و تحلیل خوشه‌ایی:

اطلاعات پایگاه داده را می‌توان به یک سری از زیرمجموعه‌های معنادار تقسیم نمود، این زیرمجموعه‌ها را به عنوان خوشه می‌شناسند. دسته‌بندی یا خوشه‌ایی کردن، درک را از واقعیت عینی افزایش می‌دهد، که پیش نیاز توصیف مفاهیم و تشخیص انحراف است. فن‌آوری خوشه‌بندی عمدتاً شامل روش‌های سنتی الگوشناسی و طبقه‌بندی ریاضی است. در اوایل سال‌های ۱۹۸۰، مچالسکی[1] مفهوم تکنولوژی خوشه‌بندی را توسعه داد و بیان نمود که نکته کلیدی در این مفهوم این است که هنگامی که اشیا دسته‌بندی می‌شوند، نه تنها فاصله‌ی بین آن‌ها درک می‌شود، بلکه نیازی برای شرح دلالت‌های ضمنی بین مفاهیم تقسیم شده احساس می‌شود که مشکل یک بعدی بودن فن‌آوری‌های سنتی را پوشش می‌دهد.

۴- توصیف مفاهیم:

دلالت ضمنی نوع معینی از اشیا را توصیف می‌کند و ویژگی‌های مرتبط آن‌ها را خلاصه می‌کند و به دسته‌ی توصیف ویژگی‌ها و توصیف تفاوت‌ها تقسیم می‌شود. توصیف ویژگی‌ها، خصوصیات مشترک اشیا را توصیف می‌کند در حالی که دسته‌ی دوم تفاوت‌های میان انواع مختلف اشیا را نمایان می‌کند. استخراج یک نوع توصیف ویژگی، تنها به مشترکات میان اشیا مربوط می‌شود.

۵- تشخیص انحراف

معمولاً در پایگاه داده ثبت‌های غیر نرمال وجود دارند، بنابراین مشخص کردن انحراف اهمیت بسیاری دارد. میزان انحراف شامل دانش بالقوه‌ی زیادی است، از آن جمله می‌توان یک مثال غیر عادی در طبقه‌بندی، یک مورد خاص که از قانون تبعیت نمی‌کند، انحراف میان مشاهدات و مدل پیش‌بینی، کمیت ارزشی‌ایی که در طول زمان تغییر می‌کند و ... را بیان نمود. روش اصلی تشخیص انحراف، یافتن تفاوت میان مشاهدات و مقادیر مبنا است.

[1] : Mchalski

داده‌کاوی معمولا با داده‌های انبوه، نامطمئن و ناقص سر و کار دارد که از این بابت شبیه به سایر رشته‌های علوم کاربردی است. حل مسائل تصمیم‌گیری در شرایط عدم اطمینان پیچیده با استفاده از مدل‌های کامل قطعی ریاضی امکان پذیر نمی‌باشد؛ حتی اگر چنین فرآیندی روی دهد، حل و تجزیه و تحلیل آن کاری بس دشوار خواهد بود. به منظور نیل به این مهم، کمک به تحلیل هوشمندانه‌ی داده‌ها و بررسی خودکار موارد متعدد، نسل جدیدی از ابزارهای محاسباتی همچون تئوری مجموعه راف، تئوری فازی، الگوریتم ژنتیک و شبکه‌های عصبی به وجود آمده است. این فن‌آوری‌های محاسباتی و ترکیب مزایای آن‌ها، به منظور رفع عدم‌دقت، عدم‌قطعیت، استدلال تقریبی و صحت جزئی در فرآیند تصمیم‌گیری به وجود آمده‌اند تا عملکردی قدرتمند، کم‌هزینه و مشابه با تصمیمات انسانی را فراهم سازند.

از آن‌جا که بسیاری از عدم قطعیت‌ها در دنیای واقعی به حدی پیچیده هستند که هیچ مدل دقیق و تحلیلی‌ایی نمی‌تواند آن‌ها را توصیف و به دقت اندازه‌گیری کند و فن‌آوری‌های سنگین کنترل شده همیشه نمی‌توانند مؤثر واقع شوند، حضور شهود انسانی به شدت احساس می‌شود که شامل درک بر مبنای تفکر انسانی است. تئوری مجموعه راف، سیستم‌های خاکستری، نظریه احتمال، مجموعه فازی، شبکه‌های عصبی، و سایر فن‌آوری‌های محاسبات نرم با استدلال در تفکر انسانی و ظرفیت فوق‌العاده نامحدود ذهن انسان در یادگیری و بهره‌گیری کامل از دانش بصری، تطابق دارند که همگی ابزارهای مؤثری در مواجه با تصمیم‌گیری در شرایط عدم قطعیت هستند. فن‌آوری محاسبات نرم در مقابل دقت، قطعیت و مطلق بودن فن‌آوری محاسبات سخت، در به‌دست آوردن راه‌حل‌هایی با دقت و سطح مطلوبیت کمتر اما اقتصادی‌تر مؤثرتر عمل می‌کنند [۷۴].

تکنیک‌های بسیاری در زمینه محاسبات نرم در داده‌کاوی به چشم می‌خورد که هر یک ویژگی‌ها، اهداف و الگوریتم خاص خود را دارد که با توجه به شرایط مسأله مورد استفاده قرار می‌گیرند. برخی از

این روش‌ها عبارت‌انداز شبکه‌های عصبی، الگوریتم ژنتیک، درخت تصمیم، ماشین برداری برتر[1]، تئوری راف ست و... .

۳-۱-۶ دلایل گزینش تئوری مجموعه راف

- ✓ تئوری مجموعه راف در برخورد با عدم قطعیت‌ها به صورت عینی و مفعولی عمل می‌کند. به بیان ساده این تئوری را می‌توان با استفاده از داده‌های خام و یا دست اول به خدمت گرفت و نیازی به داشتن اطلاعات اضافی احساس نمی‌شود. داده های جمع‌آوری شده از دنیای واقعی ممکن است حاوی انواع اختلالات، بسیاری از عدم قطعیت‌ها و اطلاعات ناقص باشند. روش‌های سنتی برخورد با عدم قطعیت مانند تئوری فازی، تئوری احتمال، تئوری شهود و... به اطلاعات اضافی یا دانش اسبق داده نیاز دارند، بنابراین مواجه شدن با حجم عظیمی از داده‌ها برای آنها دشوار خواهد بود. تفاوت عمده‌ی روش راف ست با سایر روش‌هایی که با عدم قطعیت رو به‌رو می‌شوند این است که تنها به آن دسته از داده‌هایی نیاز دارد که برای حل مسأله جمع‌آوری شده‌اند و اطلاعات اضافی را همچون تابع احتمال در آمار و یا درجه‌ی تعلق در تئوری فازی، طلب نمی‌کند. بنابراین می‌توان گفت که این تئوری در برخورد با عدم قطعیت‌ها به صورت عینی و مفعولی عمل می‌کند.
- ✓ این تکنیک توانایی تحلیل هر دو دسته معیارهای کمی و کیفی را دارد، علاوه بر آن قابلیت تلفیق راف ست با سایر روش‌ها مثل سیستم‌های خاکستری و تئوری فازی این امکان را فراهم می‌کند که از مزایای آنان نیز بهره‌مند گردد.
- ✓ مجموعه راف دارای ویژگی حذف ضایعات است و الگوریتم ساده‌سازی نسبتاً ساده‌ایی دارد. حداقل دانشی که به واسطه قوانین تصمیم‌گیری از این مدل مشتق می‌شود، می‌تواند آن را تفسیر کند.

[1] : Super Vector Machines

- ✓ این تئوری برای تناقضات و ناهماهنگی‌ها نیازی به اصلاح و ترمیم ندارد و می‌تواند قوانین ناسازگار را به دو دسته‌ی قوانین قطعی و قوانین غیرقطعی تقسیم کند.
- ✓ درک و فهم نتایج حاصل از این تکنیک، ساده و سریع است.

۳-۲ تئوری مجموعه راف[1]

داده‌کاوی و کشف دانش توجه بسیاری از دانشمندان را به خود جلب کرده است، در این میان تئوری راف ست یکی از ابزارهای بسیار موفق کشف دانش در زمینه‌های استدلال داده، تحلیل تصمیم و پیش‌بینی، کشف دانش از پایگاه داده، سیستم‌های تخصصی و سیستم‌های پشتیبانی تصمیم‌گیری محسوب می‌شود.

تئوری مجموعه راف روش ریاضیاتی نوینی است که از نیازهای کاربردی سرچشمه گرفته است و به منظور تجزیه و تحلیل داده‌ها و داده‌کاوی آن‌ها شکل گرفته است. این تکنیک که ارتباط تنگاتنگی با کشف دانش از پایگاه داده (KDD) دارد با توجه به موارد زیر روش و ابزار جدیدی برای KDD محسوب می‌شود:

۱. موضوعی که KDD در مورد آن مطالعه می‌کند، اغلب پایگاه داده‌ی رابطه‌ایی است و جدول رابطه را می‌توان به عنوان شکل خاصی از جدول تصمیم‌گیری در RST مشاهده کرد و این قضیه به کاربرد روش RST کمک می‌کند.

۲. جایگاه قوانین قطعی در دنیای واقعی واضح و مشخص است، در این میان کشف اطلاعات عدم‌قطعی از پایگاه داده، حضور RST را پر رنگ‌تر می‌کند.

۳. در فرآیند کشف دانش، مشخص کردن شرایط غیرنرمال در بین داده‌ها و استخراج نظم و قانون از درون این اختلال و آشوب، تنها با استفاده از روش‌های قدرتمندی هم‌چون راف ست امکان‌پذیر است.

[1] : Rough Set Theory : RST

۴. به کار بستن الگوریتم کشف دانش با استفاده از روش راف ست، عملکردی موازی به همراه دارد که منجر به افزایش اثربخشی فرآیند داده‌کاوی می‌شود. این امر در پایگاه‌های عظیم داده از اهمیت فراوانی برخوردار است.

۵. استفاده از روش راف ست به منظور آماده‌سازی داده‌ها و حذف ویژگی‌های مازاد، موجب افزایش اثربخشی داده‌کاوی و کاهش نرخ خطا می‌شود.

۶. قواعد تصمیم‌گیری و فرآیند استدلال روش راف ست در مقایسه با روش‌های فازی و شبکه‌ی عصبی، قابل فهم‌تر و راحت‌تر هستند [۷۵].

تئوری مجموعه راف توسط محقق لهستانی، زدیسلاو پاولاک[1] در سال ۱۹۸۲ ارائه شد. از آن‌جا که تحقیقات اولیه این تئوری بیشتر در لهستان منتشر شده بود، نظر دانشمندان علوم کامپیوتر و ریاضیات را به خود جلب نکرده و تنها در کشورهای شرق اروپا مورد استفاده قرار گرفت، اما پیش از به پایان رسیدن ده‌ی ۱۹۸۰، این تئوری توجه پژوهشگران سراسر دنیا را به خود جلب کرد و در سال ۱۹۹۲ اولین کنفرانس علمی تئوری مجموعه راف در لهستان برگزار شد. در سال ۱۹۹۵ جامعه ACM نام این تکنیک را به عنوان موضوع علمی جدید به علوم کامپیوتری اضافه کرد و در سال ۱۹۹۸ مجله‌ی بین‌المللی دانش اطلاعات[2]، نسخه‌ایی برای RST به چاپ رسانید. پس از آن تئوری مجموعه راف در علوم مختلف از جمله ریاضیات، کامپیوتر و سایر رشته‌ها مورد مطالعه و استفاده کاربردی قرار گرفته است.

تئوری مجموعه راف یک تکنیک ریاضیاتی است که به منظور رفع نیازهای کاربردی در زمینه‌های عدم‌قطعیت و ابهام به وجود آمده است. ایده‌ی اصلی این تئوری بر مبنای رابطه غیرقابل متمایزسازی[3] است. این ایده بیان می‌کند که هر شیئ با بخش خاصی از اطلاعات در ارتباط است و تنها به وسیله‌ی بخشی از آن اطلاعات قابل تشریح می‌باشد. بنابراین، اشیا با اطلاعات موجود مشابه یا یکسان، غیر قابل

[1] : Zdzislaw Pawlak
[2] : International Journal of Information Science
[3] : Indiscernibility Relation

متمایزسازی خواهند بود و به یک مجموعه تعلق پیدا می‌کنند. بلوک‌های غیرقابل متمایزسازی که با اشیای غیرقابل متمایزسازی مجموعه مرجع شکل گرفته‌اند، اعضای مجموعه‌ها یا دانه‌های دانش[۱] را تشکیل می‌دهند. این دانه‌ها مجموعه‌های بنیادی[۲] نامیده می‌شوند که بلوک‌های سازنده دانش را شکل می‌دهند و به آن‌ها کلاس‌های هم‌ارزی[۳] نیز گفته می‌شود. دانش موجود در هر یک از کلاس‌های هم‌ارزی را می‌توان با استفاده از یک رابطه غیرقابل متمایزسازی که ایده اصلی مجموعه راف را تشکیل می‌دهد، نشان داد.

بر مبنای دانش به کار گرفته شده در تقریب زنی، مجموعه مرجع به دو دسته تقسیم می‌شود. دسته اول مجموعه‌ایی از اشیای غیرقابل متمایزسازی است که مبتنی بر مجموعه‌های ویژگی مشروط است و دسته دوم مجموعه‌ایی از اعضا هستند که به دانه‌ی دانش تخمین زده شده‌اند. دانش به کار گرفته شده در تخمین زنی، مجموعه مرجع را به کلاس تصمیم‌گیری مشتق شده از مجموعه صفات تصمیم تقسیم می‌کند. بر این اساس، یک نوع دانش می‌تواند برای تخمین دیگری مورد استفاده قرار گیرد. هنگامی که این تخمین نادقیق باشد مجموعه راف حضور به عمل می‌آورد.

به بیان ساده‌تر هر اجتماع از مجموعه‌های بنیادی یک مجموعه دقیق نامیده می‌شود و سایر مجموعه‌ها که با استفاده از اجتماع مجموعه‌های بنیادین قابل توصیف نمی‌باشند، مجموعه راف می‌نامند.

نقطه شروع راف ست، مجموعه داده‌های جمع‌آوری شده می‌باشد که معمولاً به صورت یک جدول مرتب‌سازی می‌شوند و به آن سیستم اطلاعاتی[۴] یا پایگاه داده گفته می‌شود. عملگرهای اصلی در مجموعه راف تقریب‌های بالایی و پایینی[۵] هستند که از آن‌ها برای تعریف تبعیت و وابستگی کلی و وابستگی جزئی صفات در پایگاه داده استفاده می‌شود.

---

[1] : Granules of knowledge
[2] : Elementary Sets
[3] : Equivalence classes
[4] : Information System
[5] : Upper & Lower Approximation

۳-۲-۱ سیستم‌های اطلاعاتی و رابطه غیرقابل متمایزسازی

نمایش دانش در مدل مجموعه راف با سیستم‌های اطلاعاتی یا سیستم‌های نمایش دانش[۱] محقق می‌شود. فرم این سیستم‌ها به صورت جدولی دو بعدی است که ارتباط میان اشیا و ارزش صفات را نشان می‌دهد و شبیه به یک پایگاه داده‌ی رابطه‌ای است. در این جداول هر ستون نماینده یک صفت و هر سطر نماینده یک شیئ است، بنابراین هر سطر بیانگر مقادیر صفات مختلف شیئ همان سطر است. هنگامی که طبقه‌بندی با توجه به خصوصیات اشیا انجام می‌گیرد، هر مقدار از یک صفت با یک گروه از مجموعه مقادیر در ارتباط است. هنگامی که یک مقدار از هر صفت انتخاب می‌شود، توصیف یک شیئ اتفاق می‌افتد [۷۶].

در صورتی که در یک سیستم اطلاعاتی صفات و ویژگی‌ها به دو دسته‌ی ویژگی‌های شرطی (توصیفی، وضعیتی)[۲] و ویژگی تصمیم[۳] تقسیم شده باشد، آن جدول را جدول تصمیم[۴] می‌نامند.

یک سیستم اطلاعاتی در مجموعه راف با استفاده از عبارت $S = (U, A, V, F)$ به نمایش در می‌آید.

در این عبارت U نشان دهنده‌ی مجموعه‌ایی متناهی و غیرتهی از کلیه‌ی اشیای مورد نظر است و مجموعه مرجع نامیده می‌شود.

A بیان‌گر مجموعه‌ایی متناهی و غیرتهی از کلیه‌ی صفات اشیا است که به صورت $A = \{a_1, a_2, ..., a_{|A|}\}$ تعریف می‌شود.

V مجموعه کل مقادیری است که یک صفت می‌تواند در سیستم اطلاعاتی اختیار کند و دامنه‌ی آن صفت نامیده می‌شود: $V = \cup V_a \ , \ a \in A$

---

[1] : Knowledge Representation System
[2] : Condition Attribute
[3] : Decision Attribute
[4] : Decision Table

F تابع اطلاعات است و به صورت $F:U\times A\to V$ اعمال می‌شود که به ازای $\forall x\in U$ و $\forall a\in A$ مقدار هر یک از صفات مجموعه A، برای هر یک از اشیای موجود در مجموعه مرجع را نشان می‌دهد: $f(\mathrm{x},\mathrm{a})\in \mathrm{V}_a$. به بیان ساده‌تر تابع اطلاعات، هر یک از اشیای درون مجموعه مرجع را به مقدار صفت مورد نظر می‌برد و بدین ترتیب ارزش صفت هر شیئ در U را تعیین می‌کند. بنابراین هر شیئ x با استفاده از برداری نظیر $Des_{A(\mathrm{x})}=\left[f(\mathrm{x},\mathrm{a}_1),\mathrm{f}(\mathrm{x},\mathrm{a}_2),...,\mathrm{f}(\mathrm{x},\mathrm{a}_{|A|})\right]$ توصیف می‌گردد.

رابطه‌ی غیرقابل متمایزسازی I در مجموعه مرجع U، که به ازای هر زیرمجموعه غیرتهی از مجموعه A تعریف می‌شود ($\phi\neq P\subseteq A$)، به صورت زیر می‌باشد:

$$I=\{(\mathrm{x},\mathrm{y})\in \mathrm{U}\times \mathrm{U}:\mathrm{f}(\mathrm{x},\mathrm{q})=\mathrm{f}(\mathrm{y},\mathrm{q}),\forall \mathrm{q}\in \mathrm{P}\}$$ رابطه (۳-۱)

با توجه به رابطه (۳-۱) اگر $(\mathrm{x},\mathrm{y})\in \mathrm{I}$ صادق باشد، آنگاه x و y غیرقابل تمایز خواهند بود، به بیان ساده‌تر x و yنسبت به یک ویژگی خاص غیرقابل تمیز هستند و در کل دو عضوی که با یک اطلاعات توصیف می شوند، غیر قابل تمیز دادن خواهند بود. چنین تعریفی از غیرقابل متمایزسازی، یک رابطه هم‌ارزی را نشان می‌دهد. کلاس هم‌ارزی شامل شیئ x به صورت $I(\mathrm{x})$ نشان داده می‌شود. هر کلاس هم‌ارزی به عنوان یک اتم دانش در S درنظر گرفته می‌شود و اساس و بنیان مجموعه راف را تشکیل می‌دهد [۷۷].

۳-۲-۲ تقریب بالایی و تقریب پایینی

فرضیه اصلی روش راف ست این است که دانش در دل توانایی طبقه‌بندی نهفته شده است. می‌توان گفت روش راف ست چارچوب استانداردی است برای کشف حقیقت از میان داده‌های ناقص، که نتایج حاصل از آن با مدل‌های طبقه‌بندی یا قوانین به دست آمده از مجموعه‌ها تشریح می‌گردد.

تئوری راف ست دانش را به صورت مجموعه کلاس‌بندی می‌کند و آن را بخشی از مجموعه قرار می‌دهد. با توجه به دانش موجود، سه موقعیت پیش روی شیئ $x\in U$ در مورد تعلق آن به مجموعه‌ی

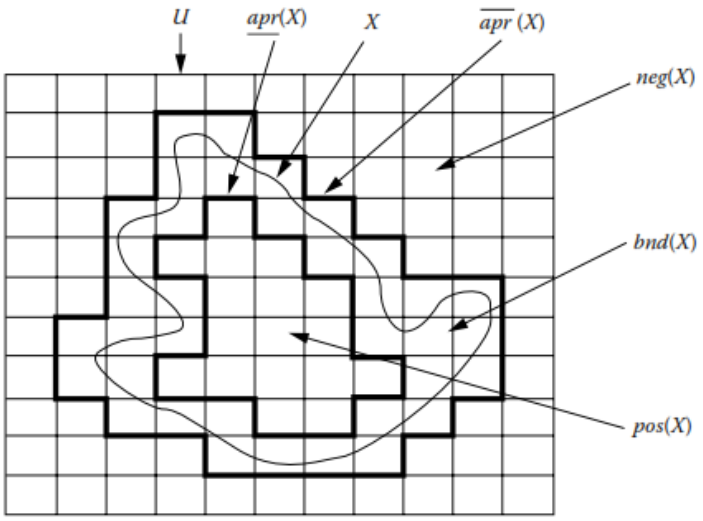
$X \subseteq U$ وجود دارد : (۱) شیئ x بی‌شک به مجموعه X تعلق دارد. (۲) شیئ x بی‌شک به مجموعه X تعلق ندارد و (۳) شیئ x ممکن است به مجموعه X تعلق داشته باشد و ممکن است تعلق نداشته باشد. با توجه به تحلیل بالا، دو مفهوم اساسی در تئوری راف ست یعنی تقریب بالایی و تقریب پایینی طراحی می‌گردد.

با در اختیار داشتن سیستم اطلاعاتی $S = (\mathrm{U}, \mathrm{A}, \mathrm{V}, \mathrm{F})$، مجموعه‌ایی غیرتهی و متناهی از صفات $P \subseteq A$، زیرمجموعه‌ایی غیرتهی و متناهی از اشیای موجود در مجموعه مرجع $X \subseteq U$ و $x \in U$ تقریب بالایی و تقریب پایینی به ترتیب با استفاده از روابط (۳-۲) و (۳-۳) تعریف می‌شوند:

رابطه (۳-۲) $$\overline{apr}_p(\mathrm{X}) = \bigcup \{x \in U : I(\mathrm{x}) \cap \mathrm{X} \neq \phi\}$$

رابطه (۳-۳) $$\underline{apr}_p(\mathrm{X}) = \bigcup \{x \in U : I(\mathrm{x}) \subseteq \mathrm{X}\}$$

تقریب پایینی مجموعه‌ی X بزرگ‌ترین پیوستگی همه‌ی اشیایی است که با اطمینان کامل به آن تعلق دارند، به این بخش از مجموعه، محدوده‌ی مثبت X گفته می‌شود و به صورت $pos_p(\mathrm{X})$ نشان داده می‌شود. همان‌طور که در شکل (۳-۴) مشخص است تقریب پایینی شامل تمامی اشیایی است که قطعاً به کلاس‌های هم‌ارزی متعلق به مجموعه X ، تعلق دارند، بنابراین بزرگ‌ترین اجتماع از کلاس‌های هم‌ارزی درون X می‌باشد.



شکل شماره (۳-۴)- طرح مفهومی مجموعه راف (تقریب پایینی و بالایی مجموعه X و نواحی مثبت و منفی آن)

تقریب بالایی مجموعه X شامل همه‌ی مجموعه‌های غیرتهی کلاس‌های هم‌ارزی است. در حقیقت کوچک‌ترین مجموعه‌ایی است که شامل اشیایی می‌شود که *ممکن است* متعلق به X باشند.

با توجه به تقریب‌های بالایی و پایینی، مجموعه U به سه منطقه مثبت، منفی و ناحیه مرزی[1] تقسیم می‌شود. منطقه مثبت به ناحیه‌ایی از مجموعه مرجع U گفته می‌شود که با توجه به صفات P، اشیای درون آن را می‌توان با اطمینان کامل متعلق به X دانست. منطقه‌ی منفی مجموعه X که با $neg_p(\mathrm{x})$ نشان داده می‌شود، مجموعه‌ایی از اشیاست که یقیناً به X تعلق ندارند، بنابراین داریم: $neg_p(\mathrm{x}) = \mathrm{U} - \overline{apr}_p(\mathrm{x})$. بنابراین می‌توان گفت که در حالت کلی عبارت $U = \overline{apr}_p(\mathrm{x}) \cup neg(\mathrm{x})$ صادق است. در این میان ناحیه مرزی مجموعه مرجع X که با $bnd_p(\mathrm{X})$ نشان داده می‌شود، تفاوت میان تقریب بالایی و تقریب پایینی در مجموعه X است که شامل اشیایی می‌شود که نمی‌توان با قطعیت اظهار کرد که آیا به X تعلق دارند یا خیر. منطقه‌ی مرزی قابل قضاوت نمی‌باشد، به عنوان مثال از نظر مجموعه ویژگی P . هیچ شیئ‌ای در این منطقه وجود ندارد که بتوان آن را به X یا $-X$ نسبت داد. در صورت تهی بودن منطقه‌ی مرزی، مجموعه X با توجه به I مجموعه‌ایی دقیق[2] خواهد بود و در صورت غیرتهی بودن این ناحیه است که مجموعه X با توجه به I مجموعه‌ی راف محسوب می‌شود [۷۸].

۳-۲-۳ وابستگی صفات[3]، دقت[4] و کیفیت تقریب[5]

وابستگی صفات:

با توجه به رابطه غیرقابل متمایزسازی، تعریف فاکتورهای مهم سیستم اطلاعاتی ساده و راحت است. یکی از مهم‌ترین این فاکتورها وابستگی صفات است. اگر تعداد کلاس‌های هم‌ارزی که از مجموعه

---

[1] : Boundary Region
[2] : Crisp Set
[3] : Attribute Dependency
[4] : Approximation Accuracy
[5] : Quality of Approximation

صفات A استخراج می‌شود، برابر باشد با تعداد کلاس‌های هم‌ارزی‌ایی که از $A-a_i$ استخراج می‌شود، آنگاه ویژگی یا صفت a_i حشو و اضافه شناخته می‌شود، در غیر این صورت ویژگی a_i در A ضروری و لازم‌الوجود خواهد بود.

با توجه به تعریف موجود از سیستم اطلاعاتی S و رابطه غیرقابل متمایزسازی I، اگر $P \subseteq A$ و $Q \subseteq A$ آنگاه وابستگی مجموعه صفات به صورت زیر خواهد بود:

۱. اگر $I(P) \subseteq I(Q)$ باشد، آنگاه مجموعه صفات Q به مجموعه صفات P وابسته خواهد بود و به صورت $P \Rightarrow Q$ نشان داده می‌شود.

۲. اگر $P \Rightarrow Q$ و $Q \Rightarrow P$، آنگاه مجموعه ویژگی P با مجموعه ویژگی Q هم‌ارز خواهد بود و به صورت $P \Leftrightarrow Q$ نشان داده می‌شود.

۳. اگر هیچ یک از دو عبارت $P \Rightarrow Q$ و $Q \Rightarrow P$ صادق نباشند، آنگاه مجموعه ویژگی P و مجموعه ویژگی Q مستقل از یکدیگر خواهند بود.

دقت تقریب:

دقت تقریب که با استفاده از رابطه (۳-۴) محاسبه می‌شود منعکس کننده‌ی درجه ادراک مجموعه X با توجه به دانش موجود است. در این رابطه که $P \subseteq A$ و $X \subseteq U$ است، علامت $|.|$ بیان‌گر تعداد اعضای مجموعه مورد نظر است.

رابطه (۳-۴)
$$\alpha_p(X) = \frac{\left|\underline{apr}_p(X)\right|}{\left|\overline{apr}_p(X)\right|}$$

دقت تقریب ضریب عددی‌ای بین صفر و یک است $0 \le \alpha_p(X) \le 1$ () که اگر برابر با یک باشد $\alpha_p(X) = 1$، آنگاه مجموعه X با توجه به I مجموعه‌ایی دقیق خواهد بود و اگر کوچکتر از یک باشد $\alpha_p(x) < 1$ آنگاه مجموعه X با توجه به I مجموعه‌ایی راف خواهد بود.

کیفیت تقریب:

ضریب عددی کیفیت تقریب یا کیفیت کلاسه‌بندی[1] به منظور سنجش میزان وابستگی دانش[2] با استفاده از رابطه (۳-۵) محاسبه می‌گردد که در این رابطه، $X = \{X_1, X_2, ..., X_n\}$ افرازی از مجموعه مرجع U است و $X_i (i = 1,2,...,n)$ یکی از کلاس‌های مجموعه X می‌باشد و داریم $P \subseteq A$.

رابطه (۳-۵)

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

کیفیت تقریب در حقیقت نشان‌دهنده نسبت تعداد اشیای به درستی کلاسه‌بندی شده به تعداد کل اشیای مجموعه مرجع است و فراوانی نسبی اشیای به درستی کلاسه‌بندی شده توسط مجموعه صفات P را نشان می‌دهد.

اگر کیفیت تخمین $\gamma_p(X) = 1$ باشد، آنگاه دانش X به صورت کامل به مجموعه صفات P وابسته خواهد بود.

اگر $0 < \gamma_p(X) < 1$ باشد، آنگاه دانش X به صورت نسبی به مجموعه صفات P وابسته خواهد بود که بیان‌گر این نکته است که تنها بخشی از صفات مجموعه P در دسترس می‌باشند و یا پایگاه داده از ابتدا نواقصی داشته است.

اگر کیفیت تخمین $\gamma_p(X) = 0$ باشد، آنگاه دانش X به صورت کامل از مجموعه صفات P مجزا خواهد بود [۷۹].

[1] : Classification Quality
[2] : Dependency of Knowledge

۳-۲-۴ محاسبه تقلیل‌ها[1] و هسته[2] با استفاده از ماتریس متمایزسازی[3]

هنگامی که از مجموعه ویژگی‌های شرطی، تعدادی از ویژگی‌ها حذف شوند، می‌توان اهمیت یک ویژگی را با محاسبه تغییرات وابستگی با استفاده از رابطه (۳-۶) ارزیابی کرد. فرض کنید $P \subseteq A$، $p \in P$ و $Q \subseteq A$، آنگاه اهمیت ویژگی $sgf\ (p,Q)$ برابر است با:

$$sgf\ (p,Q) = \gamma_P(Q) - \gamma_{P-|p|}(Q)$$ رابطه (۳-۶)

در واقع رابطه (۳-۶) اختلاف میزان وابستگی را قبل و بعد از حذف صفت می‌سنجد. هرچه میزان تغییرات وابستگی شدیدتر باشد، میزان اهمیت صفت بیشتر بوده است. از این خصیصه می‌توان به منظور حذف صفات با درجه اهمیت پایین استفاده کرد.

اگر مجموعه صفات وابسته باشد، آنگاه می‌توان به جست‌وجوی کوچک‌ترین زیرمجموعه از ویژگی‌ها پرداخت که تعداد اعضای برابری در مقایسه با مجموعه کامل صفات دارد. برای تشخیص وابسته بودن یا نبودن مجموعه ویژگی‌ها (صفات)، بررسی این‌که آیا پس از حذف هر ویژگی، تعداد اعضای مجموعه‌ها کاهش می‌یابد یا خیر، ضروری می‌باشد. هسته و تقلیل ویژگی، دو مفهوم اساسی در مجموعه‌های راف هستند. تقلیل کوچک‌ترین زیرمجموعه مستقل صفات است که با تمامی مجموعه‌های صفات، بخش اطلاعات یکسانی دارد و بخش اساسی سیستم اطلاعاتی محسوب می‌شود که می‌تواند به منظور مشخص کردن (تمیز دادن) همه اشیائی که احتمال دارد در سیستم اطلاعات اصلی قابل تشخیص باشند، مورد استفاده واقع شود. هسته، بخش مشترک همه‌ی تقلیل‌هاست.

شوارون[4] روشی برای توصیف دانش با استفاده از ماتریس متمایزسازی در سال ۱۹۹۱ ابداع نمود که این روش از مزایای بی‌شماری برخوردار است. به طور خاص، این ماتریس قادر است به راحتی هسته و تقلیل سیستم اطلاعاتی را محاسبه و توصیف نماید.

در این روش با داشتن سیستم اطلاعات $S = (U, A, V, F)$ و تعداد اشیای مجموعه مرجع $|U| = n$

---

[1] : Reducts
[2] : Core
[3] : Discernable Matrix
[4] : Showeron

ماتریس متمایزسازی S، که ماتریسی $n \times n$ است و هر عضو آن به صورت رابطه (۳-۷) تعریف می‌شود، ساخته می‌شود.

$$d(x,y) = \{a \in A | f(x,a) \neq f(y,a)\}$$ رابطه (۳-۷)

در رابطه (۳-۷)، $d(x,y)$ یک مجموعه ویژگی است که قادر است شیئ x,y را از یکدیگر تمیز بدهد.

اعضای این ماتریس توسط مجموعه‌ایی که شامل همه ویژگی‌های اعضای مجموعه متمایزسازی $[X]_i$ و $[X]_j$ است، تعریف می‌شود و به صورت d_{ij} نمایش داده می‌شود. بنابراین برای محاسبه عضو d_{ij} ویژگی‌های اعضای مجموعه $[X]_i$ و $[X]_j$ باید مشخص شوند. در حقیقت سطرها و ستون‌های این ماتریس، کلاس‌های هم‌ارزی مجموعه مورد بررسی هستند و درایه‌های این ماتریس صفاتی هستند که قادرند کلاس‌های هم‌ارزی مربوطه را از یکدیگر تفکیک نمایند. به عنوان مثال $d_{۳۱}$ درایه سطر سوم و ستون اول ماتریس متمایزسازی است که نشان‌دهنده مجموعه صفاتی است که قادراند کلاس هم‌ارزی سوم را از کلاس هم‌ارزی اول متمایز سازند.

واضح است که $d_{ij} = d_{ji}$، $d_{ij} \neq \varnothing$ و ماتریس متمایزسازی، ماتریسی متقارن می‌باشد. بنابراین ماتریس متمایزسازی را میتوان با محاسبه ماتریس پایین مثلثی به‌دست آورد.

استخراج تقلیل‌ها از ماتریس متمایزسازی توسط تابع بولی[1] یا تابع متمایزسازی[2] صورت می‌گیرد. این تابع که در رابطه (۳-۸) نشان داده شده است، بر مبنای جبر بولی[3]، تنها مقادیر صفر و یک را اختیار می‌کند و برای هر یک از ویژگی‌های مجموعه صفات $a \in A$، متغیر بولی $\alpha,''$ را در نظر می‌گیرد. اگر $a(x,y) = \{a_۱, a_۲, ..., a_k\} \neq \varnothing$، آنگاه تابع بولی برابر است با $a_۱ \vee a_۲ \vee ... a_k$ که به عنوان $\sum a(x,y); if\ a(x,y) = \varnothing$ مشخص شده است و ثابت بولی آن، یک می‌شود.

[1] : Boolean Function
[2] : Discernibility Function
[3] : Boolean Aljebra

$$f(A) = \prod_{(x,y)\in U\times U} \sum a(x,y)$$ رابطه (۳-۸)

توابع متمایزسازی به دست آمده اغلب طولانی هستند و درک آن‌ها با دشواری صورت می‌گیرد، بنابراین ساده‌سازی آنان می‌تواند به ادراک بیشتر تقلیل‌ها کمک کند. توابع متمایزسازی با استفاده از قانون جذب[1] ساده می‌شوند و بدین ترتیب تقلیل‌ها استخراج می‌گردند.

کیفیت کلاسه‌بندی مجموعه ویژگی‌های تقلیل یافته مشابه با کیفیت کلاسه‌بندی مجموعه ویژگی‌های اصلی است، بنابراین اگر حداقل مجموعه ویژگی $P \subseteq C \subseteq A$ در عبارت $\gamma_p(\mathrm{X}) = \gamma_C(\mathrm{X})$ صدق کند، آن‌گاه مجموعه P تقلیلی از مجموعه C بوده و به شکل $RED(\mathrm{P})$ نشان داده می‌شود.

تقلیل حداقل زیرمجموعه ویژگی‌های شرطی، بدون هیچ ویژگی اضافه یا حشو است که یک فرآیند کلاسه‌بندی درست را تضمین می‌کند و حداقل یک عضو مشابه با هر بخش غیرتهی از ماتریس متمایزسازی دارد. به صورت کلی تعداد زیادی تقلیل در سیستم اطلاعاتی وجود دارد، در حالی که می‌توان از یکی از آن‌ها به منظور توصیف سیستم استفاده نمود. در خلال حذف بعضی از مجموعه ویژگی‌های غیر ضروری، سیستم اطلاعاتی ساده‌سازی می‌شود و می‌توان همه مجموعه‌های ابتدایی در سیستم را شناسایی کرد.

با توجه به دشوار بودن محاسبات مربوط به تقلیل‌ها، الگوریتم‌های مختلفی برای نیل به این هدف مورد استفاده قرار گرفته است و در نرم افزارهای محاسباتی مجموعه راف مورد استفاده قرار می‌گیرند که یکی از دقیق‌ترین و سریع‌ترین آن‌ها الگوریتم ژنتیک می‌باشد.

هنگامی که بیش از یک تقلیل در سیستم اطلاعاتی وجود داشته باشد، محل تلاقی تمامی آن‌ها را هسته یا $Core$ می‌نامند. در حقیقت هسته، مجموعه‌های تک عضوی ویژگی‌ها در ماتریس متمایزسازی و بخش مشترک تمامی تقلیل‌هاست که از سیستم اطلاعاتی غیرقابل حذف می‌باشد. در واقع مجموعه تمامی صفات ضروری و لازم‌الوجود مجموعه P است که به صورت $Core(\mathrm{p})$ نمایش داده می‌شود و مهم‌ترین مجموعه، از صفات سیستم اطلاعاتی محسوب می‌شود و می‌تواند مجموعه‌ایی تهی باشد. این

[1] : Absorptivity Law

مجموعه با مقایسه تقلیل‌های حاصله و مشخص کردن بخش‌های مشترک آنان با استفاده از رابطه () به‌دست می‌آید [۸۰, ۸۱, ۸۲].

$$CORE(\mathrm{P}) = \bigcap_{R_i \in RED(\mathrm{P})} R_i, \quad (\mathrm{i} = ۱, ۲, ..., \mathrm{n}) \qquad \text{رابطه (۳-۹)}$$

۳-۲-۵ جدول تصمیم[۱] و استخراج قوانین[۲]

مفهوم وابستگی صفات[۳] می‌تواند برای توضیح ارتباط میان علت و معلول‌ها در پایگاه داده مورد استفاده قرار گیرد. با توجه به تقلیل داده‌ها، قوانین تصمیم‌گیری محتمل و تعریف شده می‌توانند توسط وابستگی کامل و وابستگی جزئی حاصل شوند.

اگر در سیستم اطلاعاتی $S = (U, A, V, f)$، ویژگی‌های A را بتوان به دو زیرمجموعه مجزا تقسیم نمود، به نحوی که یکی ویژگی‌های شرطی C و دیگری ویژگی‌های تصمیم D باشد و داشته باشیم $A = C \cup D$، $C \cap D = \varnothing$، آنگاه S را جدول تصمیم می‌نامند. نمونه‌ایی از جدول تصمیم در جدول شماره (۳-۱) نشان داده شده است.

در این جدول تصمیم تعداد n شیئ از مجموعه مرجع U و تعداد m صفت از مجموعه صفات P در نظر گرفته شده‌اند که $P \subseteq A$ می‌باشد. q_{nm} مقادیر مختلف صفات برای هر شیئ می‌باشد و d_n مقادیر ویژگی تصمیم برای هر عضو را نشان می‌دهد.

---

[1] : Decision Table
[2] : Rule Acquisition
[3] : Attribute Dependence

جدول (۳-۱)- جدول تصمیم در مجموعه راف

U	a_1	a_2	a_3	$\cdots$	a_m	D
x_1	q_{11}	q_{12}	q_{13}	$\cdots$	q_{1m}	d_1
x_2	q_{21}	q_{22}	q_{23}	$\cdots$	q_{2m}	d_2
x_3	q_{31}	q_{32}	q_{33}	$\cdots$	q_{3m}	d_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	q_{n1}	q_{n2}	q_{n3}	$\cdots$	q_{nm}	d_n

وابستگی صفات یکی از مفاهیم بسیار مهم در کاربرد جدول تصمیم می‌باشد. به طور کلی، اگر تمامی مقادیر صفت‌های D به صورت منحصر به فرد به وسیله مقادیر صفات C تعیین شوند، آنگاه D به صورت کامل به C وابسته می‌باشد و وابستگی تابعی بین مقادیر D و C وجود دارد. اگر تنها برخی از مقادیر صفات D توسط مقادیر ویژگی‌های C محاسبه شود، آنگاه وابستگی تابعی میان آنان برقرار خواهد بود. وابستگی تابعی بیانگر درجه‌ای است که ساختار دانه‌های دانش C می‌توانند ساختار دانه‌های دانش D را نمایش دهند. در تکنیک راف ست، کشف وابستگی به منظور تجزیه و تحلیل دانش، داده‌کاوی و استدلال کلی داده‌ها اهمیت بسزایی دارد.

استخراج قوانین، در هر سیستم یادگیرنده از اهمیت بالایی برخوردار است. به مجموعه تمامی متغیرهای شرطی موجود در مجموعه مرجع، کلاس‌های شرطی S گفته می‌شود و به صورت $X_i = (\mathrm{i} = 1,2,...,\mathrm{n})$ نشان داده می‌شود. به مجموعه تمامی متغیرهای تصمیم موجود در مجموعه مرجع، کلاس‌های تصمیم S گفته می‌شود و به صورت $Y_i\,(\mathrm{j} = 1,2,...,\mathrm{n})$ نشان داده می‌شود. اگر $X_i \cap Y_j = \varnothing$ ، آنگاه رابطه (۳-۱۰) برقرار خواهد بود.

رابطه (۳-۱۰) $$r : Des_C(\mathrm{X}_i) \Rightarrow \mathrm{Des}_D(\mathrm{Y}_j)$$

رابطه (۳-۱۰) که بدان قوانین تصمیم (C,D) گفته می‌شود، به صورت $\{r_{ij}\}$ نشان داده می‌شود.

اگر $X_i \subseteq Y_j$ به ازای هر $\forall i, j$ موجود در مجموعه صدق کند، آنگاه قانون r_{ij} در S قطعی[1] خواهد بود، در غیر این صورت غیرقطعی می‌باشد. به مجموعه قواعد تصمیم یک جدول تصمیم، الگوریتم تصمیم می‌گویند. در صورتی که تمامی قوانین یک الگوریتم تصمیم قطعی باشد، آن الگوریتم را سازگار و در غیر این صورت ناسازگار می‌نامند.

ساختار هر یک از قوانین به صورت رابطه (۳-۱۱) می‌باشد.

$$IF\ f(x,q_1)=r_{q_1} \wedge f(x,q_2)=r_{q_2} \wedge \ldots \wedge f(x,q_p)=r_{qp}, Then\ x \in Y_{j_1} \vee Y_{j_2} \vee \ldots Y_{jk} \quad \text{رابطه (۳-۱۱)}$$

$$Where\ \{q_1,q_2,\ldots,q_p\} \subseteq C;\ (r_{q_1},r_{q_2},\ldots,r_{qp}) \in V_{q_1} \times V_{q_2} \times \ldots \times V_{qp}.$$

به بخش *If* قوانین، قسمت شرطی و به بخش *Then* آن قسمت تصمیم می‌گویند. هر دو قسمت شرطی و تصمیم قوانین می‌تواند بیش از یک صفت شرطی و کلاس تصمیم اختیار کند [۸۳,۸۴].

۳-۲-۶ ارزیابی قواعد

به منظور ارزیابی و مقایسه قواعد به دست آمده و انتخاب قواعد مناسب‌تر، فاکتورهایی جهت سنجش ابعاد مختلف قوانین به کار گرفته می‌شود. این فاکتورها که نگاهی دقیق‌تر به قواعد دارند به چهار دسته فاکتور پشتیبان[2]، فاکتور توان[3]، فاکتور قطعیت[4] و فاکتور پوشش[5] تقسیم می‌شوند.

در صورتی که جدول تصمیم به صورت $S=(U,C,D)$ تعریف شود، $C=\{c_1,c_2,\ldots,c_n\}$ بیان‌گر متغیرهای شرطی و $D=\{d_1,d_2,\ldots,d_m\}$ بیان‌گر متغیرهای تصمیم خواهد بود و یک قاعده تصمیم در S به صورت $c_1(x),\ldots,c_n(x) \to d_1(x),\ldots,d_m(x)$ تعریف می‌شود که به اختصار با نماد $C \to D$ نشان داده می‌شود. در این صورت فاکتورهای ذکر شده به صورت زیر تعریف می‌شوند.

---

[1] : Decisive Rule
[2] : Support Factor
[3] : Strength Factor
[4] : Certainty Factor
[5] : Coverage Factor

۱- فاکتور پشتیبان:

این فاکتور بیان‌گر تعداد اشیایی است که در مجموعه از قانون مورد نظر تبعیت می‌کنند و با استفاده از رابطه (۳-۱۲) محاسبه می‌گردد:

$$Support_x(C,D) = |A(x)| = |C(x) \cap D(x)|$$ رابطه (۳-۱۲)

در این رابطه عبارت $|.|$ نشان‌دهنده‌ی تعداد اعضای آن مجموعه می‌باشد. هر چه فاکتور پشتیبان یک قانون بیشتر باشد، میزان اطمینان به آن قانون افزایش می‌یابد.

۲- فاکتور توان:

این فاکتور نسبت دفعات تکرار قانون مورد نظر (تعداد اشیای موجود در مجموعه که از این قانون تبعیت می‌کنند) به کل اشیای موجود در مجموعه مرجع است و با استفاده از رابطه (۳-۱۳) محاسبه می‌گردد.

$$Strength_x(C,D) = \sigma_x(C,D) = \frac{Support_x(C,D)}{|U|}$$ رابطه (۳-۱۳)

۳- فاکتور قطعیت:

این فاکتور که نشان‌دهنده‌ی احتمال شرطی تعلق y به $D(x)$، به شرط متعلق بودن آن به $C(x)$ است، با استفاده از رابطه (۳-۱۴) محاسبه می‌گردد.

$$Cer_x(C,D) = \frac{|C(x) \cap D(x)|}{|C(x)|} = \frac{Supp_x(C,D)}{|C(x)|} = \frac{\sigma_x(C,D)}{\pi[C(x)]}$$
$$Where\ \pi[C(x)] = \frac{|C(x)|}{|U|}$$ رابطه (۳-۱۴)

این فاکتور میزان دقیق بودن قانون مورد نظر را می‌سنجد و مقدار آن در نرم‌افزار روزتا[1]، تحت عنوان "دقت[2]" محاسبه می‌گردد.

اگر $Cer_x(C,D)=1$ برقرار باشد، آنگاه قانون مورد نظر را قانون تصمیم قطعی[3] می‌نامند و اگر $0<Cer_x(C,D)<1$ صادق باشد، آنگاه قانون مورد نظر را قانون تصمیم غیرقطعی[4] می‌نامند.

۴- فاکتور پوشش:

این فاکتور میزان عمومیت داشتن یک قانون در جدول تصمیم را بیان می‌کند و با استفاده از رابطه (۳-۱۵) محاسبه می‌شود.

$$Coverage_x(C,D)=\frac{|C(x)\cap D(x)|}{|D(x)|}=\frac{Supprot_x(C,D)}{|D(x)|}=\frac{\sigma_x(C,D)}{\pi[D(x)]}$$

$$Where\ D(x)\neq\phi \text{ and } \pi[D(X)]=\frac{|D(x)|}{|U|}$$

رابطه (۳-۱۵)

[1] : Rosetta
[2] : Accuracy
[3] : Certain Decision Rule
[4] : Uncertain Decision Rule

فصل چهارم

تجزیه و تحلیل داده‌ها

۴-۱ مقدمه

با توجه به اهمیت بازده سهام و نقدشوندگی آن، تحلیل رفتار سهام، بررسی واکنش‌های بازده سهام نسبت به متغیرهای کلان اقتصادی و استخراج قواعد نظام‌مند مکنون حاکم بر آن، تأثیر بسزایی در تصمیم‌گیری سرمایه‌گذاران خواهد داشت. از این رو به منظور نیل به این هدف و بررسی فرضیات و سؤالات موجود، در این مرحله از تحقیق ابزار و روش‌های جمع‌آوری اطلاعات، جامعه آماری و سهام انتخاب شده برای انجام پژوهش و متغیرهای شرطی و تصمیم در نظر گرفته شده در جدول تصمیم، به تفصیل بیان شده‌اند. گردآوری و آماده‌سازی داده‌ها یا استفاده از نرم‌افزار EXCEL صورت گرفته و پس از آن فرآیند تجزیه و تحلیل داده‌های گردآوری شده با استفاده از تکنیک راف ست و نرم‌افزار ROSETTA انجام شده و نتایج حاصل از این تحقیق که در قالب قوانین استخراجی از پایگاه‌های داده مربوطه به نمایش در می‌آید، توصیف شده است.

۴-۲ جامعه آماری و نمونه

جامعه آماری در این پژوهش شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌باشد، اما با توجه به شرایط مدیریتی، استراتژیک، تأمین مالی و گروه صنعتی که هر شرکت دارد، رفتار بازده سهام و نقدشوندگی آن نسبت به سایر شرکت‌ها متفاوت خواهد بود، بنابراین به منظور تجزیه و تحلیل سؤالات پژوهش و استخراج الگوهای نظام‌مند حاکم بر رفتار بازده سهام، می‌بایستی شرکتی به عنوان نمونه انتخاب گردد. با استفاده از نرم‌افزار ره‌آوردنوین و بررسی سهام شرکت‌های مختلف، شرکت سایپا که زیرمجموعه‌ی صنعت خودروسازی در بورس اوراق بهادار می‌باشد، به دلیل دارا بودن تعداد روزهای معاملاتی بالا در طی سال‌های مورد بررسی (۱۳۸۸-۱۳۹۲) انتخاب شده است.

خودروسازی ایران دومین شاخه‌ی مهم اقتصاد ملی، بعد از صنعت نفت و گاز می‌باشد. در سال ۲۰۰۹ کشور ایران به عنوان دوازدهمین تولیدکننده بزرگ جهانی خودرو شناخته شد و موفق شد از سال ۲۰۰۸ صادرات گسترده محصولات خودروسازی خود را به ۳۰ کشور جهان آغاز کند. در حال حاضر ۱۲ واحد

تولیدی در زمینه‌ی خودروسازی فعالیت می‌کنند که مهم‌ترین آن‌ها شرکت سایپا و ایران‌خودرو هستند.

شرکت سایپا در سال ۱۳۴۶ در زمینی با مساحت ۲۴۰ هزار متر مربع و زیربنای ۲۰ هزار متر مربع (در حال حاضر فقط مساحت زمین کارخانه مرکزی ۴۱۵ هزار متر مربع می‌باشد) با سرمایه اولیه ۱۶۰ میلیون ریال به نام " شرکت سهامی تولید اتومبیل سیتروئن ایران" تاسیس گردید. در تاریخ ۱۵ اسفند ۱۳۴۵ ثبت و در اواخر سال ۱۳۴۷ به مرحله بهره‌برداری رسید.

این شرکت تولید اولین محصولات خود را که شامل "وانت آکا" و سواری "ژیان" بود با روش کاملاً دستی و بدون بهره‌گیری از تجهیزات و امکانات مدرن آغاز نمود. تولیدات شرکت بعد از سال ۱۳۵۳ به واسطه استفاده از ابزارهای جدید و مکانیزه شدن برخی از بخش‌های تولیدی، سیر صعودی یافت و بر تنوع محصولات شرکت نیز افزوده شد. به‌طوری که علاوه بر تولید ژیان با مدل‌های مهاری، پیکاب، از سال ۱۳۵۵ تولید رنو در مدل‌های ۲ درب و ۴ درب شروع و تا سال ۱۳۷۲ ادامه داشت.

در اوایل سال ۱۳۵۴ نام شرکت با حذف کلمه سیتروئن از انتهای عبارت فرانسوی آن به "شرکت سهامی ایرانی تولید اتومبیل" به نام اختصاری (سایپا SAIPA) که مأخوذ از عبارت فرانسوی ‘‘Societe Annonyme Iraniane De Production Automobile’’ می‌باشد، تغییر یافت.

شرکت در ۱۶ تیرماه ۱۳۵۸ تحت مالکیت دولت جمهوری اسلامی ایران در آمده و از ۱۸ آذر ماه ۱۳۶۰ تحت سرپرستی سازمان گسترش و نوسازی صنایع ایران قرار گرفت و در تاریخ ۱۳۶۵/۲/۱ بر اساس مصوبه هیات وزیران، کلیه سهام سرمایه آن به نمایندگی از طرف دولت جمهوری اسلامی به نام سازمان گسترش و نوسازی صنایع ایران منتقل گردید.

در سال ۱۳۷۷ با توجه به سیاست‌های جدید سازمان گسترش و نوسازی صنایع ایران مبنی بر تمرکز بیشتر در فعالیت‌های صنعت خودروسازی کشور، استفاده بهینه از امکانات موجود در مجموعه شرکت‌های تحت پوشش سازمان، ایجاد رقابت و ... وزارت صنایع زمینه‌های ادغام مجموعه‌های خودروسازی را در قالب دو گروه بزرگ خودروساز ایجاد کرد و بدین ترتیب گروه خودروسازی سایپا و گروه خودروسازی ایران خودرو شکل گرفتند. گروه خودروسازی سایپا به منظور تولید انواع خودروها خصوصاً تولید

خودروهای سنگین و تجاری در همان سال با خرید ۸۰ درصد از سهام شرکت سایپا دیزل (ایران کاوه) و ۶۳/۲ درصد از سهام زامیاد و در سال ۱۳۷۸ با خرید ۵۱ درصد از سهام شرکت پارس خودرو گام‌های عمده ای در این زمینه برداشت.

در دی ماه سال ۱۳۷۸ به پیروی از سیاست‌های دولت جمهوری اسلامی ایران مبنی بر کاهش تصدی دولت و خصوصی سازی شرکت‌های دولتی و به موجب تبصره ۳۵ قانون بودجه کل کشور، بیش از ۵۱٪ سهام این شرکت واگذار شد و بدین ترتیب سایپا نیز در زمره شرکت‌های خصوصی قرار گرفت. امروزه شرکت سایپا با در اختیار داشتن بیش از ۹۰ شرکت خودروساز، قطعه ساز، تحقیقاتی، مالی و رفاهی به صورت تابعه و وابسته (مستقیم و غیر مستقیم)، به یک گروه خودروسازی بزرگ با امکان تولید " انواع مختلف خودروهای سواری و تجاری " تبدیل شده به طوری که در عرصه جهانی به عنوان بیست و یکمین شرکت خودروساز بین‌المللی مطرح شده است.

۳-۴ متغیرهای تحقیق

با توجه به مطالب ذکر شده در فصل سوم، تکنیک مجموعه راف نیازمند دو دسته متغیر می‌باشد. دسته اول متغیرهای شرطی و دسته دوم متغیر تصمیم می‌باشد.

با توجه به اجرای فرآیند راف ست در دو مرحله‌ی بررسی رفتار بازده سهام به صورت مجزا و بررسی رفتار بازده سهام با استفاده از موج برگشتی نقدشوندگی، به دو مجموعه جدول تصمیم نیاز می‌باشد.

جدول تصمیم بررسی رفتار بازده سهام به صورت مجزا، دارای متغیر تصمیم ۱) بازده سهام شرکت سایپا و متغیرهای شرطی ۲) شاخص نرخ بهای کالاها و خدمات مصرفی (نرخ تورم)، ۳) قیمت جهانی نفت خام، ۴) قیمت جهانی طلا، ۵) متوسط قیمت فروش دلار آمریکا در بازار آزاد شهر تهران، ۶) متوسط قیمت فروش یک سکه تمام بهار آزادی (طرح قدیم) در بازار آزاد شهر تهران، ۷) پایه پولی ۸) شاخص قیمت مسکن (سال پایه ۱۳۹۰) ۹) شاخص قیمت بورس اوراق بهادار تهران (TEPIX) و ۱۰) شاخص بازده نقدی و قیمت بورس اوراق بهادار تهران (TEDPIX) ۱۱) شاخص صنعت خودروسازی می‌باشند.

بررسی رفتار بازده سهام با استفاده از موج برگشتی نقدشوندگی، نیازمند جدول تصمیم متفاوتی است که در آن متغیر تصمیم معیار تعدیل‌شده‌ی عدم نقدشوندگی آمیهیود با حجم صفر می‌باشد و متغیرهای شرطی در این جدول همان ده متغیر شرطی ذکر شده برای جدول تصمیم بازده سهام به اضافه سری زمانی بازده سهام شرکت سایپا می‌باشند. بنابراین دوازده متغیر تحقیق در این پژوهش بکار گرفته شده است که در ادامه به تفصیل توضیح داده شده‌اند.

۴-۳-۱ توصیف متغیرهای تحقیق

۱) بازده سهام شرکت سایپا:

مقادیر این متغیر که متغیر تصمیم این پژوهش می‌باشد، با استفاده از نرم‌افزار ره‌آورد نوین به صورت ماهانه طی سال‌های مورد بررسی (۱۳۸۸-۱۳۹۲) گردآوری شده است. بازده شرکت سایپا در هر ماه بیان‌گر بازده تجمعی روزهای معاملاتی آن ماه می‌باشد.

۲) شاخص بهای کالاها و خدمات مصرفی (نرخ تورم):

شاخص بهای کالاها و خدمات مصرفی یکی از انواع شاخص‌های قیمت است که معیار سنجش تغییرات قیمت مقدار ثابت و معینی از کالاها و خدمات مصرفی خانوارهای شهرنشین بوده و وسیله‌ایی است برای اندازه‌گیری سطح عمومی قیمت‌ها در بازار خرده فروشی و یکی از بهترین ابزارهای سنجش نرخ تورم و قدرت خرید پول کشور می‌باشد. اداره آمار اقتصادی بانک مرکزی متصدی محاسبه شاخص بهای کالاها و خدمات مصرفی در ایران است، بنابراین داده‌های مورد نیاز در این زمینه به صورت ماهانه و از بخش شاخص‌های قیمت سایت بانک مرکزی جمهوری اسلامی ایران به آدرس www.cbi.ir و در بازه‌ی زمانی مورد بررسی جمع‌آوری شده است.

۳) قیمت جهانی نفت خام:

قیمت جهانی نفت خام که به صورت قیمت هر بشکه نفت خام به دلار آمریکا در نظر گرفته شده است، با استفاده از سایت مدیریت اطلاعات انرژی آمریکا به آدرس www.eia.gov و در بازه‌ی زمانی مورد نظر جمع‌آوری شده است.

۴) قیمت جهانی طلا:

قیمت جهانی طلا که به صورت قیمت هر انس طلا به دلار آمریکا در نظر گرفته شده است، با استفاده از سایت شورای جهانی طلا به آدرس www.gold.org و در بازه‌ی زمانی مورد نظر گردآوری شده است.

۵) متوسط قیمت فروش دلار آمریکا در بازار آزاد شهر تهران:

متوسط قیمت فروش دلار آمریکا در بازار آزاد شهر تهران که به عنوان یکی از نرخ‌های ارز رایج توسط بانک مرکزی جمهوری اسلامی ایران بر حسب ریال اعلام می‌گردد، به صورت ماهانه و از بخش شاخص‌های قیمت سایت بانک مرکزی جمهوری اسلامی ایران به آدرس www.cbi.ir و در بازه‌ی زمانی مورد بررسی جمع‌آوری شده است.

۶) متوسط قیمت فروش یک سکه تمام بهار آزادی (طرح قدیم) در بازار آزاد شهر تهران:

داده‌های مربوط به این شاخص که به صورت ماهانه توسط بانک مرکزی جمهوری اسلامی ایران اعلام می‌گردد نیز از بخش شاخص‌های قیمت سایت بانک مرکزی جمهوری اسلامی ایران به آدرس www.cbi.ir و در بازه‌ی زمانی مورد بررسی جمع‌آوری شده است.

۷) پایه پولی :

پایه پولی به نسبت میان اسکناس عرضه شده توسط بانک مرکزی و مجموع دارایی‌های این بانک که به عنوان پشتوانه اسکناس، قابل قبول باشد گفته می‌شود. هرچه این نسبت بزرگتر باشد، میزان نقدینگی و به تبع آن تورم بیشتر است. به عبارت دیگر بزرگ بودن این نسبت نشانگر پایین بودن ارزش پول است. منابع پایه پولی همان دارایی‌های بانک مرکزی است که شامل:۱ -ذخایر ارزی ۲- استقراض دولت از بانک مرکزی ۳- سایر دارایی‌ها و ۴-تسهیلات اعطایی به بانکهای تجاری می‌شود. مصارف پایه پولی همان بدهی‌های پولی می‌باشد که شامل: ۱- نسبت ذخایر قانونی (RR) ۲-ذخایر اضافی بانک‌ها نزد بانک مرکزی و ۳-نسبت اسکناس و مسکوکات در دست مردم می‌باشد. داده‌های مرتبط با این متغیر به صورت ماهانه از سایت بانک مرکزی جمهوری اسلامی ایران به آدرس www.cbi.ir و در بازه‌ی زمانی مورد بررسی جمع‌آوری شده است.

۸) شاخص قیمت مسکن (سال پایه ۱۳۹۰)

منظور از شاخص‌های مسکن، اطلاعات مناسب عملی درباره جنبه‌های مختلف اقتصادی، اجتماعی، فرهنگی، زیست محیطی و کالبدی مسئله مسکن است. شاخص‌های مسکن به سیاست‌گزاران کلان بخش مسکن کمک می‌کنند تا تصویر روشن‌تری از شرایط مسکن درگذشته، حال و آینده به دست آورند و سیاست‌ها و راهبردهای متناسبی اتخاذ نمایند.از این رو این شاخص‌ها از یک سو ابزار شناخت وضعیت مسکن در ابعاد مختلف هستند و از سوی دیگر، ابزار کلیدی برای ترسیم چشم انداز آینده مسکن و برنامه ریزی آن محسوب می‌شوند. به همین دلیل می‌توان شاخص‌های مسکن را کلیدی‌ترین ابزار برنامه‌ریزی و شالوده اصلی آن دانست.

شاخص‌های مسکن در ایران تحت تاثیر دو نوع عوامل بیرونی و درونی قرار داشته‌اند. عوامل بیرونی موثر بر مسکن، نظیر عوامل جمعیتی، اقتصادی و اجتماعی، موجبات رونق و رکود مسکن را فراهم کرده‌اند و برنامه‌های مسکن کمتر در وضعیت کلی مسکن و بهبود شاخص های آن موثر بوده اند. عوامل درونی

مسکن، که عمدتا" تحت تاثیر چگونگی و فرآیند برنامه ریزی مسکن قرار دارند نیز در ابعاد خاصی، وضعیت مسکن را شکل می دهند. داده‌های مرتبط با این متغیر به صورت ماهانه از سایت بانک مرکزی جمهوری اسلامی ایران به آدرس www.cbi.ir و در بازه‌ی زمانی مورد بررسی جمع‌آوری شده است.

۹) شاخص قیمت بورس اوراق بهادار تهران (TEPIX):

شاخص کمیتی است که نماینده چند متغیر همگن می‌باشد و وسیله‌ایی برای اندازه‌گیری و مقایسه پدیده‌هایی است که دارای ماهیت و خاصیت مشخصی هستند که بر مبنای آن می‌توان تغییرات ایجاد شده در متغیرهای معینی را در طول یک دوره بررسی نمود.

شاخص‌ها نماگرهایی هستند که وظیفه عیان ساختن شرایط عمومی قیمت یا بازدهی در میان تمام یا گروهی از شرکت‌های پذیرفته شده در بورس را بر عهده گرفته‌اند. از این رو می‌توان شاخص را نماگری نامید که نشان دهنده‌ی سطح عمومی پارامتری مشخص (به طور معمول قیمت یا بازدهی) در میان گروهی از متغیرهای مورد بررسی (تمام یا گروهی از شرکت‌ها) است. بنابراین شاخص کل قیمت نماگری است که سطح عمومی قیمت کل شرکت‌های پذیرفته شده در بورس را نشان می‌دهد.

شاخص کل قیمت (TEPIX) که مخفف کلمه Tehran Price Index می‌باشد یکی از اصلی‌ترین شاخص‌های بورس اوراق بهادار تهران است که روش محاسبه آن بر اساس فرمول لاسپیرز می‌باشد.

فرمول لاسپیرز:

{(ارزش جاری سهم در بازار*تعداد سهام منتشره) تقسیم بر(ارزش پایه سهام*تعداد سهام در سال پایه)}ضرب در عدد ۱۰۰

داده‌های مربوط به این متغیر با استفاده از نرم‌افزار ره‌آوردنوین و طی دوره‌ی مورد نظر به صورت ماهانه جمع‌آوری شده است.

۱۰)شاخص قیمت و بازده نقدی (TEDPIX):

این شاخص نشان‌دهنده‌ی تغییرات قیمتی و بازده نقدی سهام در بورس اوراق بهادار تهران می‌باشد، به عبارتی نشان‌دهنده تغییرات بازدهی کل سرمایه‌گذاری در بورس اوراق بهادار می‌باشد و به وسیله آن می‌توان فرصت سرمایه‌گذاری در بورس اوراق بهادار را با دیگر بازارها مقایسه نمود. داده‌های مربوط به این شاخص نیز با استفاده از نرم‌افزار ره‌آورد نوین در طی سال‌های مورد نظر جمع‌آوری شده است.

۱۱) شاخص صنعت خودروسازی

فرمول محاسبه این شاخص همانند محاسبه شاخص کل می‌باشد و نشان دهنده‌ی تغییرات در ارزش سهام شرکت‌های صنعتی می‌باشد و برای هر صنعت بطور جداگانه محاسبه می‌گردد. اطلاعات مربوط به این شاخص از مرکز پردازش اطلاعات مالی ایران (شرکت مدیریت فناوری بورس تهران) به آدرس fipiran.com (Financial Information Processing of Iran) در بازه‌ی زمانی مورد بررسی جمع‌آوری شده است.

۱۲) متغیر تصمیم نقدشوندگی:

معیار تعدیل‌شده‌ی عدم‌نقدشوندگی آمیهیود[1] با حجم صفر (AdjILLIQ) به عنوان متغیر تصمیم نقدشوندگی در این پژوهش در نظر گرفته شده است. این معیار که لگاریتم طبیعی میانگین نسبت بازده روزانه سهام به حجم ریالی معامله شده در هر روز از آن ماه، ضرب در یک به اضافه درصد روزهای حجم صفر سهام در آن ماه می‌باشد با استفاده از رابطه (۲-۶) و با اطلاعات جمع‌آوری شده از نرم‌افزار ره‌آورد نوین محاسبه گردیده است.

[1] : Modified Version of the Amihud Illiquidity Measure

۴-۴ مطالعه موردی و تجزیه و تحلیل داده‌ها

مراحل پاکسازی و یکپارچه‌سازی کشف دانش با استخراج داده‌های مورد نیاز از پایگاه‌های داده‌ی موجود (سایت بانک مرکزی جمهوری اسلامی ایران، اطلاعات بورس اوراق بهادار تهران، شورای جهانی طلا و سایت مدیریت اطلاعات انرژی آمریکا) و ساختن انبار داده صورت پذیرفته است. مراحل انتخاب و تبدیل کردن داده‌ها با استخراج داده‌های مرتبط با سال‌های مورد بررسی از انبار داده و تبدیل کردن آن‌ها به نوع داده‌های مورد نیاز برای تحلیل، انجام گرفته و داده‌های منتخب شناسایی شده‌اند. زین پس مراحل مرتبط با داده‌کاوی آغاز می‌گردد. (به منظور بررسی مراحل مختلف کشف دانش به شکل شماره (۳-۱) رجوع شود).

یکی از مهم‌ترین و ابتدایی‌ترین مراحل داده‌کاوی، آماده‌سازی داده‌های منتخب می‌باشد. این آماده‌سازی می‌تواند با توجه به نوع و شرایط داده‌های منتخب شامل تکنیک‌های پالایش داده، الگوریتم‌های تکمیل جداول اطلاعاتی ناقص، تحلیل داده‌های خارج از محدوده، دسته‌بندی، خوشه‌بندی و کاهش داده‌ها باشد. آماده‌سازی داده‌ها متکی به نوع ماهیت داده و نوع تحلیل آن‌ها می‌باشد، بنابراین با شناخت داده‌ها می‌توان روش مناسب برای آماده‌سازی آن‌ها انتخاب نمود.

داده‌های بکارگرفته شده در این پژوهش از نوع داده‌های کمی پیوسته پویا می‌باشد. هرگاه صفت خاصه و داده‌ی مورد نظر را بتوان شمارش و یا اندازه‌گیری نمود و سپس آن را به صورت عدد بیان کرد، آن را یک متغیر کمی یا عددی می‌نامند که دارای دو خاصیت ترتیب و فاصله بوده و به دو دسته متغیرهای پیوسته[1] و گسسته[2] تقسیم‌بندی می‌شوند.

متغیرهای پیوسته، متغیرهایی هستند که می‌توانند کلیه‌ی مقادیر حقیقی بین محدوده‌ای را داشته باشند. نوعی از این داده‌ها که مقدارشان با زمان تغییر نمی‌کند را داده‌های ایستا و نوعی که با گذر زمان تغییر می‌کنند، داده‌های پویا می‌نامند.

---

[1] : Continuous
[2] : Discrete

۴-۴-۱ آماده‌سازی داده‌ها

آماده‌سازی داده‌ها با در نظر گرفتن نوع و ویژگی‌های داده‌های مورد بررسی در این پژوهش، از نوع کاهش داده می‌باشد. کاهش داده‌ها خود به سه دسته‌ی ۱) کاهش صفات خاصه ۲) کاهش نمونه‌ها و ۳) کاهش مقادیر یک صفت خاصه تقسیم می‌شود که در این پژوهش فرآیند کاهش مقادیر یک صفت خاصه می‌بایستی اعمال گردد.

۴-۴-۱-۱ کاهش مقادیر یک صفت خاصه

مقادیر موجود برای یک صفت خاصه محدوده‌ایی را تشکیل می‌دهند که بدون شک این مقادیر زیرمجموعه‌ایی از دامنه‌ی صفت خاصه مزبور است. هدف اصلی تکنیک‌های کاهش مقادیر صفات خاصه، کاهش تعداد آن‌ها می‌باشد. تکنیک‌های گسسته‌سازی یا مجزاسازی[۱] می‌توانند برای کاهش مقادیر در پایگاه داده استفاده شوند. وظیفه‌ی این تکنیک‌ها قرار دادن داده‌های پیوسته در تعدادی فاصله می‌باشد. این تکنیک‌ها عموماً در توصیف ساده‌تر داده‌ها و فهم راحت‌تر و در نهایت نتایج پایانی فرآیند داده‌کاوی تأثیر بسزایی دارند. برخی از این تکنیک‌ها می‌توانند توسط انسان به صورت دستی یا در تعریف شمای پایگاه داده‌ها تعریف شوند. تکنیک‌های دسته‌بندی، روش آماری کای دو، گسسته‌سازی مبتنی بر آنتروپی و .... روش‌هایی از این دست هستند[۸۵, ۸۶, ۸۷, ۸۸].

در نرم‌افزار ROSETTA امکان بهره‌گیری از نه تکنیک گسسته‌سازی زیر وجود دارد:

۱) گسسته‌ساز (مقیاس‌گر) استدلال بولی متعامد[۲]

۲) گسسته‌ساز دستی[۳]

۳) گسسته‌ساز مبتنی بر آنتروپی[۴]

۴) گسسته‌ساز مبتنی بر گروه‌بندی فراوانی یکسان[۱]

---

[1] : Discretization Methods
[2] : Boolean Reasoning Orthogonal Scaler (BROrthogonal Scaler)
[3] : Manual Scaler
[4] : Entropy Scaler (Entropy/MDL algorithm)

۵) گسسته‌ساز ساده[۲]

۶) گسسته‌ساز تقریباً ساده[۳]

۷) گسسته‌ساز ستونی متعامد[۴]

۸) فرم RSES گسسته‌ساز استدلال بولی متعامد

۹) فرم RSES گسسته‌ساز ستونی متعامد

در این پژوهش با توجه به خصوصیات داده‌های شرطی و تصمیم از تکنیک‌های گسسته‌سازی ۱) استدلال بولی متعامد، ۲) تکنیک مبتنی بر آنتروپی و ۳) تکنیک مبتنی بر گروه‌بندی فراوانی یکسان به صورت مجزا استفاده شده است.

۱) گسسته‌ساز استدلال بولی متعامد

این تکنیک که توسط نوین و اسکورون[۵] معرفی شده است مقاطع حاصل از روش گسسته‌ساز Naive را به خدمت می‌گیرد، روش فوق تعداد بسیار زیادی مقطع ایجاد می‌کند که حتی ممکن است بیش از حد نیاز باشد. تکنیک گسسته‌ساز استدلال بولی متعامد مقاطع حاصل از روش Naïve را با استفاده از استدلال بولی ساده‌سازی می‌نماید و نهایتاً زیرمجموعه باقی‌مانده، حداقل مجموعه گسسته‌سازی است که متمایزسازی سیستم اطلاعاتی را به عهده دارد. این الگوریتم در ابتدا یک تابع بولی به نام f که در رابطه (۴-۱) نشان داده شده است را با استفاده از مقاطع کاندید تشکیل داده و سپس عملکرد اولیه‌ی آن را مورد سنجش قرار می‌دهد[۸۹].

$$f = \prod_{(x,y)} \sum_{a} \left\{ \sum c^* \middle| c \in C_a \; and \; a(x) < c < a(y) \text{ and } \partial_A(x) \neq \partial_A(y) \right\} \qquad \text{رابطه (۴-۱)}$$

نتایج حاصل از گسسته‌سازی داده‌ها با استفاده از روش گسسته‌ساز استدلال بولی متعامد در جدول شماره (۴-۳) نشان داده شده است. در جدول شماره (۴-۱) مقادیر حداقل و حداکثر ده شاخص مورد استفاده و

---

[1] : Equal Frequency Scaler (Equal Frequency Bining)
[2] : Naïve Scaler (Naive algorithm)
[3] : SemiNaive Scaler (Semi-Naive algorithm)
[4] : Orthogonal File Scaler (From File with Cuts)
[5] : Nguyen and Skowron

در جدول شماره (۴-۲) مقادیر حداقل و حداکثر تغییرات ده شاخص مورد استفاده نشان داده شده است.

جدول شماره (۴-۱)- حداقل و حداکثر ده شاخص مورد استفاده

	نرخ تورم	قیمت نفت	قیمت طلا	نرخ دلار	قیمت سکه
Min	۷۰	۵۰/۶۸	۹۰۰/۶۸	۹۸۰۸	۲۱۶۳۹۱۹
Max	۱۸۶/۶	۱۰۷/۸	۱۸۱۵/۱۴	۳۶۹۹۴	۱۴۴۹۵۹۰۰
	پایه پولی	قیمت مسکن	TEPIX	TEDPIX	شاخص صنعت
Min	۴۸۲۷۹۷/۳	۷۱/۹	۸۱۲۲	-۶/۷۲	۲۶۴۰/۹۸
Max	۱۱۴۹۴۳۵/۸	۱۲۸/۹	۸۷۳۹۶	۱۷/۶۴	۱۴۲۲۳/۷۱

جدول شماره (۴-۲)- حداقل و حداکثر تغییرات ده شاخص مورد استفاده

	نرخ تورم	قیمت نفت	قیمت طلا	نرخ دلار	قیمت سکه
Min	۰	-۱۳/۷۱	-۹/۲۶	-۹/۴۶	-۱۵/۲۷
Max	۵/۳۳	۲۵/۲۹	۱۰/۶۱	۳۴/۹۴	۳۷/۷۶
	پایه پولی	قیمت مسکن	TEPIX	TEDPIX	شاخص صنعت
Min	-۱۱/۹۴	-۸/۳۰	-۶/۶۲	-۱۴/۹۱	-۱۴/۹۳
Max	۱۶/۷۶	۱۰/۹۹	۱۶/۱۳	۱۳/۵۵	۳۳/۵۵

جدول شماره (۴-۳)- گسسته‌سازی داده‌های تغییرات ده شاخص با استفاده از روش استدلال بولی متعامد

نرخ تورم	قیمت نفت	قیمت طلا	نرخ دلار	قیمت سکه
[*,۱/۰۶۸۲۶)	[*,۰/۶۹۷۲۳)	No Discretization	No Discretization	[*,۰/۴۸۱۱۵)
[۱/۰۶۸۲۶,*)	[۰/۶۹۷۲۳,*)	No Discretization	No Discretization	[۰/۴۸۱۱۵,*)
پایه پولی	قیمت مسکن	TEPIX	TEDPIX	شاخص صنعت
[*,۱/۱۱۹۶۸)	[*,۰/۴۵۵۵۹)	[*,۰/۷۰۹۰۷)	[*,-۱/۰۲۳۶۹)	No Discretization
[۱/۱۱۹۶۸,*)	[۰/۴۵۵۵۹,*)	[۰/۷۰۹۰۷,*)	[-۱/۰۲۳۶۹,*)	No Discretization

۲) تکنیک مبتنی بر آنتروپی

آنتروپی یکی از معیارهای رایج جهت گسسته‌سازی به شمار می‌رود. گسسته‌سازی مبتنی بر آنتروپی یک تکنیک جداسازی بالا به پایین است، که بر روی داده‌های با برچسب کلاس انجام می‌شود. این بدین معنی است که روش با کمک اطلاعات توزیع کلاس، نقاط انفصال را برای داده‌ها محاسبه می‌کند. نقاط انفصال آن‌هایی هستند که مقادیر داده‌ها بر مبنای آن‌ها افراز می‌شوند. برای گسسته‌سازی صفت خاصه‌ایی مانند A، روش مزبور مقداری از A را به عنوان نقطه‌ی انفصال انتخاب می‌کند که مقدار آنتروپی را حداقل کند و این کار به صورت بازگشتی جهت یافتن نقاط انفصال دیگر ادامه می‌یابد. چنان‌چه مجموعه داده‌های D برای هر یک از نمونه‌ها دارای برچسب کلاس باشند، روش گسسته‌سازی صفت خاصه‌ی پیوسته‌ی A شامل سه مرحله است.

هر مقدار صفت خاصه‌ی A می‌تواند بالقوه یک نقطه‌ی انفصال باشد. پس از مرتب‌سازی مقادیر A به بررسی تک تک حالت‌ها می‌پردازیم. با انتخاب یک نقطه انفصال SP، مجموعه داده‌های D به دو زیرمجموعه‌ی $D_۱$ و $D_۲$ شکسته می‌شود که به ترتیب هر یک از این زیرمجموعه‌ها شروط $A \le SP$ و $A > SP$ را ارضا می‌کنند.

برای هر یک از نقاط انفصال تعریف شده در مرحله‌ی قبل رابطه (۴-۲) محاسبه می‌شود.

$$Info_A(\mathrm{D}) = \frac{|D_۱|}{|D|} \times Entropy(\mathrm{D}_۱) + \frac{|D_۲|}{|D|} \times Entropy(\mathrm{D}_۲)$$ رابطه (۴-۲)

که در آن $|D|$ تعداد نمونه‌ها در مجموعه‌ی D را نشان می‌دهد و برای آنتروپی رابطه (۴-۳) بر قرار است.

$$Entropy(\mathrm{D}_۱) = -\sum_{i=۱}^{k} p_i \times \log_۲(\mathrm{p}_i)$$ رابطه (۴-۳)

که در آن p_i احتمال کلاس C_i در $D_۱$ را نشان می‌دهد. در نهایت از میان نقاط انفصال، نقطه‌ای انتخاب می‌شود که مقدار $Info$ محاسبه شده‌ی آن کوچکترین باشد. به عبارت دیگر این نقطه، داده‌ها را با درجه‌ی خلوص بهتری طبقه‌بندی می‌کند، یعنی تنوع کلاس در هر طبقه به حداقل

می‌رسد. نتایج حاصل از گسسته‌سازی داده‌ها با استفاده از روش آنتروپی در جداول شماره (۴-۴)، (۴-۵) و (۴-۶) نشان داده شده است[۹۰،۹۱].

جدول شماره (۴-۴)- گسسته‌سازی داده‌های تغییرات ده شاخص با استفاده از روش آنتروپی

نرخ تورم	قیمت نفت	قیمت طلا	نرخ دلار
[*, 0.42375)	[*, -5.33526)	[*, -4.36183)	[*, -4.00989)
[0.42375, 0.46729)	[-5.33526, -5.16496)	[-4.36183, -4.04230)	[-4.00989, -2.02186)
[0.46729, 0.52216)	[-5.16496, -5.10950)	[-4.04230, -3.40644)	[-2.02186, -0.61488)
[0.52216, 0.60797)	[-5.10950, -5.04559)	[-3.40644, -2.52742)	[-0.61488, -0.24994)
[0.60797, 0.95352)	[-5.04559, -4.82638)	[-2.52742, -2.38241)	[-0.24994, -0.09010)
[0.95352, 1.01580)	[-4.82638, -4.20942)	[-2.38241, -2.04225)	[-0.09010, 0.10349)
[1.01580, 1.03165)	[-4.20942, -3.48473)	[-2.04225, -1.71096)	[0.10349, 0.24047)
[1.03165, 1.06826)	[-3.48473, -2.80417)	[-1.71096, -0.70429)	[0.24047, 0.43129)
[1.06826, 1.15687)	[-2.80417, -2.68589)	[-0.70429, -0.60928)	[0.43129, 0.60516)
[1.15687, 1.16124)	[-2.68589, 0.69723)	[-0.60928, -0.49305)	[0.60516, 0.70152)
[1.16124, 1.17698)	[0.69723, 1.86380)	[-0.49305, -0.35769)	[0.70152, 0.90939)
[1.17698, 1.23457)	[1.86380, 2.54677)	[-0.35769, -0.32577)	[0.90939, 1.35227)
[1.23457, 1.25628)	[2.54677, 2.88715)	[-0.32577, 1.57707)	[1.35227, 1.69039)
[1.25628, 1.27778)	[2.88715, 3.06859)	[1.57707, 2.02741)	[1.69039, 1.96677)
[1.27778, 1.35166)	[3.06859, 3.53482)	[2.02741, 3.00480)	[1.96677, 2.12086)
[1.35166, 1.44918)	[3.53482, 3.88122)	[3.00480, 3.45107)	[2.12086, 3.35883)
[1.44918, 1.49188)	[3.88122, 4.22905)	[3.45107, 3.80218)	[3.35883, 4.14401)
[1.49188, 1.53596)	[4.22905, 4.38880)	[3.80218, 4.26637)	[4.14401, 4.63132)
[1.53596, 1.65355)	[4.38880, 5.49489)	[4.26637, 4.86093)	[4.63132, 5.70498)
[1.65355, 1.81568)	[5.49489, 6.53220)	[4.86093, 5.08447)	[5.70498, 7.54969)
[1.81568, 1.85450)	[6.53220, 6.86848)	[5.08447, 5.22840)	[7.54969, 12.47450)
[1.85450, 1.86377)	[6.86848, 7.32451)	[5.22840, 5.42412)	[12.47450, 13.06740)
[1.86377, 2.06801)	[7.32451, 7.55741)	[5.42412, 5.68560)	[13.06740, *)
[2.06801, 2.32105)	[7.55741, 7.58909)	[5.68560, 9.79819)	
[2.32105, 2.44274)	[7.58909, 8.23103)	[9.79819, *)	
[2.44274, 2.57957)	[8.23103, 11.27390)		
[2.57957, 2.65128)	[11.27390, 14.66570)		
[2.65128, 3.11235)	[14.66570, *)		
[3.11235, 3.33005)			
[3.33005, 3.40345)			
[3.40345, 4.22563)			
[4.22563, 5.11254)			
[5.11254, *)			

جدول شماره (۴-۵)- گسسته‌سازی داده‌های تغییرات ده شاخص با استفاده از روش آنتروپی

قیمت سکه	پایه پولی	قیمت مسکن	TEPIX
[*, -5.50813)	[*, -1.78992)	[*, -0.10167)	[*, -3.55362)
[-5.50813, -2.34889)	[-1.78992, -1.40717)	[-0.10167, 0.10325)	[-3.55362, -2.28690)
[-2.34889, -1.95423)	[-1.40717, -1.14754)	[0.10325, 0.23205)	[-2.28690, -0.91099)
[-1.95423, -1.52087)	[-1.14754, -0.46744)	[0.23205, 0.45559)	[-0.91099, 0.70907)
[-1.52087, -1.30286)	[-0.46744, -0.29595)	[0.45559, 0.70157)	[0.70907, 1.23022)
[-1.30286, -1.03013)	[-0.29595, -0.19200)	[0.70157, 0.81212)	[1.23022, 1.64515)
[-1.03013, -0.36717)	[-0.19200, -0.01979)	[0.81212, 0.88519)	[1.64515, 2.09897)
[-0.36717, -0.01997)	[-0.01979, 0.19501)	[0.88519, 0.98061)	[2.09897, 2.29286)
[-0.01997, 0.48115)	[0.19501, 0.27482)	[0.98061, 1.20018)	[2.29286, 2.48288)
[0.48115, 0.54378)	[0.27482, 0.47927)	[1.20018, 1.23712)	[2.48288, 3.15258)
[0.54378, 1.50411)	[0.47927, 0.79738)	[1.23712, 1.26485)	[3.15258, 3.82238)
[1.50411, 1.98690)	[0.79738, 0.94606)	[1.26485, 1.41115)	[3.82238, 4.41431)
[1.98690, 2.08674)	[0.94606, 1.11968)	[1.41115, 1.59264)	[4.41431, 4.77219)
[2.08674, 2.26226)	[1.11968, 1.18162)	[1.59264, 1.79145)	[4.77219, 4.93413)
[2.26226, 2.41822)	[1.18162, 1.27678)	[1.79145, 1.80299)	[4.93413, 5.25432)
[2.41822, 4.41479)	[1.27678, 1.60218)	[1.80299, 1.96974)	[5.25432, 5.36947)
[4.41479, 5.30034)	[1.60218, 1.89638)	[1.96974, 2.22202)	[5.36947, 7.13857)
[5.30034, 5.56035)	[1.89638, 2.38494)	[2.22202, 2.34347)	[7.13857, 7.94029)
[5.56035, 5.58997)	[2.38494, 2.52701)	[2.34347, 2.41447)	[7.94029, 8.46138)
[5.58997, 5.69313)	[2.52701, 2.70989)	[2.41447, 2.60383)	[8.46138, 10.08890)
[5.69313, 6.42439)	[2.70989, 2.78111)	[2.60383, 3.06460)	[10.08890, 10.33420)
[6.42439, 6.70111)	[2.78111, 4.44954)	[3.06460, 3.70558)	[10.33420, 10.79980)
[6.70111, 8.94809)	[4.44954, 4.48576)	[3.70558, 4.00022)	[10.79980, 11.78980)
[8.94809, 10.24310)	[4.48576, 4.71156)	[4.00022, 5.44991)	[11.78980, 12.28840)
[10.24310, 11.99150)	[4.71156, *)	[5.44991, *)	[12.28840, 13.54660)
[11.99150, 15.22340)			[13.54660, 14.87530)
[15.22340, *)			[14.87530, 16.05420)
			[16.05420, *)

جدول شماره (۴-۶)- گسسته‌سازی داده‌های تغییرات ده شاخص با استفاده از روش آنتروپی

TEDPIX		شاخص صنعت	
[*, -3.40683)	[3.05074, 3.93393)	[*, -5.71580)	[0.78354, 1.21173)
[-3.40683, -2.46210)	[3.93393, 4.56251)	[-5.71580, -5.50760)	[1.21173, 2.96787)
[-2.46210, -1.56321)	[4.56251, 4.71996)	[-5.50760, -4.27522)	[2.96787, 4.04487)
[-1.56321, -1.02369)	[4.71996, 4.76866)	[-4.27522, -3.44129)	[4.04487, 7.46923)
[-1.02369, -0.71859)	[7.93398, 8.43018)	[-3.44129, -3.15148)	[7.46923, 9.77633)
[-0.71859, -0.30578)	[8.43018, 11.96510)	[-3.15148, -1.60238)	[9.77633, 13.54390)
[-0.30578, 0.13904)	[11.96510, *)	[-1.60238, -1.51240)	[13.54390, 15.53120)
[0.13904, 0.38090)		[-1.51240, -1.45769)	[15.53120, 15.96380)
[0.38090, 0.55486)		[-1.45769, -1.36977)	[15.96380, 16.40480)
[0.55486, 0.70880)		[-1.36977, -0.56390)	[16.40480, 16.75910)
[0.70880, 0.95192)		[-0.56390, -0.41511)	[16.75910, 19.83260)
[0.95192, 1.19192)		[-0.41511, -0.26814)	[19.83260, 31.08540)
[1.19192, 2.19499)		[-0.26814, 0.34355)	[31.08540, *)
[2.19499, 3.05074)		[0.34355, 0.78354)	

۳) تکنیک مبتنی بر گروه‌بندی فراوانی یکسان

این الگوریتم روشی غیر نظارتی[1] و تک متغیره[2] می‌باشد که دسته‌بندی‌ها را با توجه به فراوانی آن‌ها انجام می‌دهد. در این روش کاربر تعداد بازه‌های مورد نظر (n) را مشخص کرده و با توجه به آن تعداد $n-1$ برش به نحوی که تقریباً در هر n بازه تعداد مساوی از اشیا قرار داشته باشند، صورت می‌گیرد.

در این پژوهش تعداد بازه‌ها برابر با چهار در نظر گرفته شده است و نتایج حاصل از گسسته‌سازی داده‌ها با استفاده از این روش در جدول (۴-۷) نشان داده شده است.

[1] : Unsupervised
[2] : Univariate

جدول شماره (۴-۷)- گسسته‌سازی داده‌های تغییرات ده شاخص با استفاده از روش گروه‌بندی فراوانی یکسان

نرخ تورم	قیمت نفت	قیمت طلا	نرخ دلار	قیمت سکه
[*, 0.98840)	[*, -4.39659)	[*, -2.52742)	[*, -0.61488)	[*, -1.30286)
[0.98840, 1.35166)	[-4.39659, 1.86380)	[-2.52742, 0.84397)	[-0.61488, 0.43129)	[-1.30286, 1.88848)
[1.35166, 2.20337)	[1.86380, 5.49489)	[0.84397, 4.26637)	[0.43129, 2.35680)	[1.88848, 5.30034)
[2.20337, *)	[5.49489, *)	[4.26637, *)	[2.35680, *)	[5.30034, *)
پایه پولی	قیمت مسکن	TEPIX	TEDPIX	شاخص صنعت
[*, -0.29595)	[*, 0.11647)	[*, 0.03481)	[*, -3.82799)	[*, -3.75264)
[-0.29595, 1.18162)	[0.11647, 0.93419)	[0.03481, 3.15258)	[-3.82799, 0.38090)	[-3.75264, 0.27456)
[1.18162, 2.70989)	[0.93419, 1.80299)	[3.15258, 7.94029)	[0.38090, 4.76866)	[0.27456, 7.98635)
[2.70989, *)	[1.80299, *)	[7.94029, *)	[4.76866, *)	[7.98635, *)

۴-۴-۲ اجرای تکنیک مجموعه راف

۴-۴-۲-۱ تشکیل سیستم اطلاعاتی

نخستین گام در اجرای تکنیک مجموعه راف تشکیل سیستم اطلاعاتی یا جدول تصمیم می‌باشد. هر سطر از این جدول مربوط به یکی از ماه‌های مورد بررسی در طول دوره‌ی پنج ساله ۸۸-۹۲ است که از فروردین ۸۸ آغاز شده و تا اسفند ۹۲، تعداد ۶۰ رکورد را تشکیل داده است. ستون‌های جدول تصمیم به دو دسته‌ی متغیرهای شرطی و متغیر تصمیم تقسیم می‌شود. متغیرهای شرطی عبارت‌اند از: نرخ بهای کالاها و خدمات مصرفی (CPI)، قیمت نفت خام، قیمت جهانی طلا، نرخ دلار در بازار آزاد، قیمت سکه بهار آزادی، پایه پولی، شاخص قیمت مسکن، شاخص کل بورس اوراق بهادار تهران (TEPIX)، شاخص قیمت و بازده نقدی بورس اوراق بهادار (TEDPIX) و شاخص صنعت خودروسازی و متغیر تصمیم بازده سهام شرکت سایپا می‌باشد. جدول تصمیم به شرح موارد ذکر شده در جدول شماره (۴-۸) قابل مشاهده می‌باشد. جدول تصمیم مربوط به فاکتور نقدشوندگی با استفاده از متغیرهای شرطی فوق‌الذکر به اضافه بازده سهام شرکت سایپا و متغیر تصمیم نقدشوندگی تشکیل شده است که در جدول شماره (۴-۹) قابل مشاهده می‌باشد.

جدول شماره (۴-۸)- جدول تصمیم با متغیر تصمیم بازده سهام

	CPI	قیمت نفت	قیمت طلا	نرخ دلار	قیمت سکه	پایه پولی	شاخص مسکن	TEPIX	TEDPIX	شاخص صنعت	بازده سهام
فرو ۸۸	$q_{۱۱}$	$q_{۱۲}$	$q_{۱۳}$	$q_{۱۴}$	$q_{۱۵}$	$q_{۱۶}$	$q_{۱۷}$	$q_{۱۸}$	$q_{۱۹}$	$q_{۱۱۰}$	$d_{۱}$
اردی۸۸	$q_{۲۱}$	$q_{۲۲}$	$q_{۲۳}$	$q_{۲۴}$	$q_{۲۵}$	$q_{۲۶}$	$q_{۲۷}$	$q_{۲۸}$	$q_{۲۹}$	$q_{۲۱۰}$	$d_{۲}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
اسف ۹۲	$q_{۶۰۱}$	$q_{۶۰۲}$	$q_{۶۰۳}$	$q_{۶۰۴}$	$q_{۶۰۵}$	$q_{۶۰۶}$	$q_{۶۰۷}$	$q_{۶۰۸}$	$q_{۶۰۹}$	$q_{۶۰۱۰}$	$d_{۶۰}$

جدول شماره (۴-۹)- جدول تصمیم با متغیر تصمیم نقدشوندگی

	CPI	قیمت نفت	قیمت طلا	نرخ دلار	قیمت سکه	پایه پولی	شاخص مسکن	TEPIX	TEDPIX	شاخص صنعت	بازده سهام	نقد شوندگی
فرو ۸۸	$q_{۱۱}$	$q_{۱۲}$	$q_{۱۳}$	$q_{۱۴}$	$q_{۱۵}$	$q_{۱۶}$	$q_{۱۷}$	$q_{۱۸}$	$q_{۱۹}$	$q_{۱۱۰}$	$q_{۱۱۱}$	$d_{۱}$
اردی۸۸	$q_{۲۱}$	$q_{۲۲}$	$q_{۲۳}$	$q_{۲۴}$	$q_{۲۵}$	$q_{۲۶}$	$q_{۲۷}$	$q_{۲۸}$	$q_{۲۹}$	$q_{۲۱۰}$	$q_{۲۱۱}$	$d_{۲}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		⋮
اسف ۹۲	$q_{۶۰۱}$	$q_{۶۰۲}$	$q_{۶۰۳}$	$q_{۶۰۴}$	$q_{۶۰۵}$	$q_{۶۰۶}$	$q_{۶۰۷}$	$q_{۶۰۸}$	$q_{۶۰۹}$	$q_{۶۰۱۰}$	$q_{۶۰۱۱}$	$d_{۶۰}$

۴-۴-۲-۲ کلاس‌های هم‌ارزی و تقریب بالایی و پایینی

عملگرهای اصلی در مجموعه راف تقریب‌های بالایی و پایینی هستند که از آن‌ها برای تعریف تبعیت و وابستگی کلی و وابستگی جزئی صفات در پایگاه داده استفاده می‌شود. علاوه بر آن تشکیل کلاس‌های هم‌ارزی نیز با توجه به تقریب‌های بالا و پایین صورت می‌گیرد. به منظور تشکیل کلاس‌های هم‌ارزی می‌توان هر یک از کلاس‌های تصمیم مورد نظر را مبنای طبقه‌بندی قرار داد. در صورتی که تنها دو کلاس

تصمیم موجود باشد، مجموعه تقریب پایین یا منطقه‌ی مثبت یک کلاس با منطقه‌ی منفی کلاس دیگر قرینه خواهد بود، بنابراین محاسبه کلاس‌های هم‌ارزی تنها برای یکی از کلاس‌های تصمیم کفایت می‌کند. در این‌جا با توجه به این که جدول تصمیم بازده سهام سایپا تنها دو کلاس تصمیم ۱) روند صعودی بازده سهام و ۲) روند نزولی بازده سهام دارد، کلاس‌های هم‌ارزی، تقریب بالایی و تقریب پایینی با توجه به روند صعودی بازده سهام محاسبه شده است. نتایج حاصل از محاسبات در جدول شماره‌ی (۴-۱۱) قابل مشاهده می‌باشد. ماه‌های مورد بررسی به ترتیب تاریخ در جدول تصمیم شماره‌گذاری شده‌اند و کلاس‌های هم‌ارزی و تقریب‌های بالایی و پایینی با استفاده از این شماره‌گذاری‌ها محاسبه گردیده‌اند.

مقدار نقدشوندگی سهام در ماه‌های مختلف که با استفاده از رابطه (۲-۶) محاسبه شده است، به منظور ساده‌سازی تحلیل به صورت فازی به نمایش در آمده‌اند. متغیرهای کلامی به کار گرفته شده، بازه‌های نقدشوندگی و اعداد نسبت داده شده به هر بازه در جدول شماره (۴-۱۰) نشان داده شده است.

جدول شماره (۴-۱۰)-متغیرهای کلامی برای نقدشوندگی

متغیر کلامی	بازه‌ی نقدشوندگی	عدد نسبت داده شده
خیلی خیلی کم	(۳۵/۱۵-۳۲/۱۵)	۵
خیلی کم	(۳۲/۱۵-۲۹/۱۵)	۴
متوسط	(۲۹/۱۵-۲۶/۱۵)	۳
خوب	(۲۶/۱۵-۲۳/۱۵)	۲
خیلی خوب	(۲۳/۱۵-۲۰/۱۵)	۱

محاسبه کلاس‌های هم‌ارزی، تقریب بالایی و تقریب پایینی برای جدول تصمیم با متغیر تصمیم نقدشوندگی در جدول شماره (۴-۱۲) نشان داده شده است. در این جدول کلاس‌های هم‌ارزی با توجه به کلاس تصمیم (۱) یا به عبارتی نقدشوندگی با متغیر کلامی "خیلی خوب" صورت پذیرفته است.

جدول شماره (۴-۱۱)- کلاس‌های هم‌ارزی، تقریب بالایی، تقریب پایینی، منطقه مرزی و منطقه منفی با متغیر تصمیم بازده سهام

کلاس‌های هم‌ارزی		تقریب بالایی		تقریب پایینی	منطقه مرزی	منطقه منفی
{۴۸ ,۳۸}	{۱۳ ,۲}	{۵۵ ,۳۷}	{۱۴ ,۸}	{۵۲}	{۴۸ ,۳۸}	{۴۵}
{۵۲}	{۵۹ ,۲۰}	{۲۳}	{۴۲}	{۵۶}	{۵۱ ,۱۷}	{۵۷}
{۵۶}	{۲۴}	{۳۱}	{۵}	{۴۹}	{۴۷ ,۴۰}	{۱}
{۴۹}	{۵۴}	{۵۱ ,۱۷}	{۴۳}	{۵۵ ,۳۷}	{۳۹ ,۱۰}	{۳۶}
{۴۵}	{۷}	{۴۷ ,۴۰}	{۵۴}	{۲۳}	{۴۶ ,۴۴ ,۱۱}	{۵۰}
{۵۵ ,۳۷}	{۶ ,۳}	{۳۹ ,۱۰}		{۳۱}	{۵۸ ,۴}	{۳۳}
{۵۷}	{۲۱}	{۴۶ ,۴۴ ,۱۱}		{۳۴}	{۵۹ ,۲۰}	{۲۲ ,۱۶}
{۲۳}	{۱۲}	{۵۸ ,۴}		{۲۴}	{۶ ,۳}	{۲۸}
{۳۰ ,۲۹ ,۱۵}	{۲۷}	{۳۴}		{۵۴}	{۳۰ ,۲۹ ,۱۵}	{۵۳}
{۳۶}	{۳۲}	{۵۹ ,۲۰}		{۲۷}		{۱۳ ,۲}
{۳۱}	{۹}	{۲۴}		{۳۲}		{۷}
{۵۰}	{۴۱}	{۳۰ ,۲۹ ,۱۵}		{۹}		{۲۱}
{۵۱ ,۱۷}	{۶۰}	{۶ ,۳}		{۳۵}		{۱۲}
{۴۷ ,۴۰}	{۳۵}	{۲۷}		{۲۶ ,۱۹}		{۴۱}
{۳۳}	{۲۶ ,۱۹}	{۳۲}		{۱۴ ,۸}		{۶۰}
{۳۹ ,۱۰}	{۱۴ ,۸}	{۹}		{۴۲}		{۱۸}
{۴۶ ,۴۴ ,۱۱}	{۱۸}	{۳۵}		{۵}		{۲۵}
{۵۸ ,۴}	{۴۲}	{۲۶ ,۱۹}		{۴۳}		
{۲۲ ,۱۶}	{۵}	{۴۸ ,۳۸}				
{۳۴}	{۲۵}	{۵۲}				
{۲۸}	{۴۳}	{۵۶}				
{۵۳}	{۱}	{۴۹}				

جدول شماره (۴-۱۲)- کلاس‌های هم‌ارزی، تقریب بالایی، تقریب پایینی، منطقه مرزی و منطقه منفی با متغیر تصمیم نقدشوندگی

کلاس‌های هم‌ارزی			تقریب بالای		تقریب پایینی		منطقه مرزی	منطقه منفی
{۳۸}	{۳۴}	{۲۷}	{۱۰}	{۴۷}	{۱۰}	{۵۰}	{۱۳ ,۲}	{۴۸}
{۴۸}	{۲۸}	{۳۲}	{۱۱}	{۳۹}	{۱۱}	{۲۳}	{۲۶ ,۱۹}	{۵۶}
{۵۲}	{۵۳}	{۹}	{۳۴}	{۳۳}	{۳۴}	{۳۱}	{۲۹ ,۱۵}	{۱}
{۵۶}	{۵۹}	{۴۱}	{۲ ,۱۳}	{۵۱}	{۸ ,۱۴}	{۵۹}		{۳۶}
{۴۹}	{۲۰}	{۶۰}	{۵۹}	{۵۰}	{۵۵ ,۳۷}	{۵۷}		{۴۴ ,۴۶}
{۴۵}	{۲۴}	{۳۵}	{۲۰}	{۲۳}	{۲۰}	{۴۵}		{۱۷}
{۵۷}	{۵۴}	{۱۸}	{۵۴}	{۳۱}	{۵۴}	{۴۹}		{۴۰}
{۲۳}	{۷}	{۴۲}	{۱۲}	{۹}	{۷}	{۵۲}		{۱۶ ,۲۲}
{۱}	{۳}	{۵}	{۷}	{۵۷}	{۲۱}	{۳۸}		{۵۸}
{۳۶}	{۶}	{۲۵}	{۲۱}	{۴۵}	{۱۲}			{۴}
{۳۱}	{۲۱}	{۴۳}	{۵۵ ,۳۷}	{۴۹}	{۲۷}			{۲۸}
{۵۰}	{۱۲}	{۳۰}	{۳۲}	{۵۲}	{۳۲}			{۵۳}
{۱۷}	{۵۵ ,۳۷}		{۲۷}	{۳۸}	{۹}			{۲۴}
{۵۱}	{۴۴ ,۴۶}		{۴۱}		{۴۱}			{۳}
{۴۰}	{۱۶ ,۲۲}		{۳۵}		{۳۵}			{۶}
{۴۷}	{۲ ,۱۳}		{۱۹ ,۲۶}		{۴۲}			{۶۰}
{۳۳}	{۱۹ ,۲۶}		{۸ ,۱۴}		{۱۸}			{۵}
{۳۹}	{۸ ,۱۴}		{۱۸}		{۴۳}			{۳۰}
{۱۰}	{۱۵ ,۲۹}		{۴۲}		{۲۵}			
{۱۱}			{۱۵ ,۲۹}		{۳۹}			
{۵۸}			{۲۵}		{۳۳}			
{۴}			{۴۳}		{۵۱}			

۴-۴-۲-۳ محاسبه تقلیل‌ها و استخراج قوانین

روش‌های بسیاری از جمله الگوریتم ژنتیک[۱]، الگوریتم 1R هولت[۲]، الگوریتم جانسون[۳]، تقلیل‌های پویا[۴] و ... به منظور محاسبه‌ی تقلیل‌ها و استخراج قوانین در تکنیک راف ست مورد استفاده قرار می‌گیرد که در این پژوهش از الگوریتم ژنتیک و هولت استفاده شده است.

الگوریتم ژنتیک

الگوریتم‌های ژنتیک که بر اساس مکانیسم علم ژنتیک طبیعت بنا نهاده شده است، به دلیل توانایی در حل مسائل بهینه‌سازی پیچیده با فضای جست‌وجوی ناشناخته مورد توجه بسیار قرار گرفته است. این الگوریتم در سال ۱۹۷۰ توسط هلند، دی‌جونگ و گلدبرگ[۵] ارائه شده و جزء کلاس الگوریتم‌های بهینه‌سازی تصادفی محسوب می‌شود. فرآیند کلی این الگوریتم‌ها به ترتیبی است که در ابتدا کار خود را با جمعیت اولیه‌ایی از نمونه‌ها به صورت تصادفی آغاز می‌کند. هر یک از نمونه‌های موجود در جمعیت نشان‌دهنده‌ی یک جواب بالقوه مسئله مورد بحث است. نمونه‌ها طی تکرارهای متوالی به نام نسل تکامل یافته و هر نسل با معیارهایی ارزیابی می‌شود. نسل‌های بعدی توسط عملگرهای ژنتیکی (مانند جهش[۶]) ساخته شده و فرآیند تکرار تا رسیدن به شرط پایانی ادامه می‌یابد. پس از گذشت تعدادی نسل، الگوریتم به سمت بهترین راه‌حل همگرا می‌شود. روش ساختن جمعیت اولیه، فرآیند ارزیابی هر نسل، عملگرهای ژنتیکی و تنظیم پارامترهای مختلف در الگوریتم ژنتیک مواردی هستند که باید به جزییات به آن‌ها پرداخته شود.

الگوریتم‌های ژنتیک به جای اینکه عملیاتی را بر روی راه‌حل‌ها انجام دهند، با یک شکل کدگذاری شده از

---

[1] : Genetic Algorithm
[2] : Holte's 1R
[3] : Johnsons' Algorithm
[4] : Dynamic Reducts
[5] : Holland, DeJong & Goldberg
[6] : Mutation

آن‌ها درگیر هستند. بدین ترتیب و با تعیین یک تابع ارزشیابی[1] مناسب در واقع اصول اولیه‌ی الگوریتم ژنتیک جهت اجرا آماده شده‌اند. اگر صفت خاصه‌ایی دارای بیش از دو مقدار باشد، تعداد بیت‌های بیشتری نیز برای کدگذاری نیاز است [۹۲].

فواید استفاده از این روش بسیار زیاد است که از آن جمله می‌توان به مزایای ۱) در این الگوریتم جمعیتی از نقاط به صورت موازی (به جای یک نقطه) مورد جست‌وجو قرار می‌گیرند ۲) از قواعد انتقال احتمالی (به جای قواعد انتقال قطعی) استفاده می‌شود ۳) در این الگوریتم نیازی به اطلاع درباره مشتق‌پذیری تابع هدف نیست و تعریف یک تابع برازش برای مسأله کفایت می‌کند ۴) به راحتی برای مسائل چندهدفه مورد استفاده قرار می‌گیرد و ۵) برای محیط‌های نویزی مفید است، اشاره نمود.

الگوریتم 1R هولت

رابرت هولت[2] که محققی در زمینه‌ی علوم کامپیوتری از دانشگاه اوتاوا[3] کانادا است در سال ۱۹۹۳ الگوریتمی به منظور استخراج قوانین بر مبنای تنها یک ویژگی ارائه نمود که دقت آن برابر با دقت قوانینی است که توسط سایر روش‌های یادگیری ماشین استخراج می‌شود.

در این الگوریتم فرض می‌شود که داده‌ها گسسته هستند، بنابراین در صورتی که داده‌ها پیوسته باشند، باید ابتدا گسسته‌سازی شوند و سپس الگوریتم اجرا شود. این تکنیک با داده‌های مفقوده به صورت یک مقدار یا ارزش مجزا برخورد می‌کند.

این الگوریتم در ابتدا ویژگی‌ایی را که دارای بیشترین بار اطلاعاتی است انتخاب کرده و آن را مبنای استخراج قوانین قرار می‌دهد و برای هر ویژگی مانند a بدین شرح قوانین را استخراج می‌کند:

برای هر مقدار v که در دامنه‌ی صفت a جای دارد، مجموعه‌ایی از اعضا که در آن a دارای مقدار v است تشکیل می‌دهد. کلاسی که بیشترین میزان تکرار را در این مجموعه دارا باشد، c می‌نامد و قانونی به شرح رابطه (۴-۴) ارائه می‌دهد.

---

[1] : Fitness Function
[2] : Robert C. Holte
[3] : Ottawa

رابطه (۴-۴) $if\ a\ has\ value\ v\ ,then\ the\ class\ is\ c.$

و پس از استخراج قانون، میزان دقت آن را محاسبه می‌نماید. این الگوریتم نشان می‌دهد که بسیاری از پایگاه‌های داده از برخی قوانین ساده پیروی می‌کنند که از دقت بالایی برخوردارند و لزومی به کاوش و جست‌وجو برای قوانین بسیار پیچیده وجود ندارد.

قوانین استخراج شده

محاسبه‌ی کاهش‌ها با استفاده از دو الگوریتم ژنتیک و هولت1R منجر به استخراج قوانین از پایگاه داده می‌شود. به منظور سنجش و مقایسه قوانین، از دو معیار دقت (قطعیت) و پوشش استفاده می‌شود که توسط نرم‌افزار ROSETTA محاسبه می‌گردد. قوانینی که از میزان دقت (قطعیت) و پوشش بیشتری برخوردارند، قابل اطمینان‌تر بوده و از میان مجموعه قوانین انتخاب می‌گردند. از میان کل قوانین به دست آمده در این پژوهش، به برخی از مهم‌ترین آن‌ها در جداول شماره‌ی (۴-۱۳)، (۴-۱۴)، (۴-۱۵) و (۴-۱۶) اشاره شده است.

در این جداول عبارت LHS Support اشاره دارد به تعداد اشیایی که در مجموعه آموزشی (در این پژوهش کل پایگاه داده) از بخش IF آن قانون تبعیت می‌کنند.

عبارت RHS Support اشاره دارد به تعداد اشیایی که در مجموعه آموزشی (در این پژوهش کل پایگاه داده) هم از بخش IF و هم از بخش Then آن قانون تبعیت می‌کنند. به بیان ساده‌تر این عدد نشان‌دهنده همان فاکتور پشتیبان یا Support است که با استفاده از رابطه (۳-۱۲) به‌دست می‌آید.

عبارت RHS Accuracy از تقسیم RHS Support بر LHS Support به دست می‌آید و نشان‌دهنده میزان فاکتور قطعیت یا همان دقت (Accuracy) است که با استفاده از رابطه (۳-۱۴) محاسبه می‌گردد.

عبارت RHS Coverage از تقسیم RHS Support بر تعداد اشیای موجود در کلاس تصمیم که در بخش Then قانون صدق می‌کنند، به دست می‌آید و نشان‌دهنده میزان فاکتور پوشش یا همان Coverage است که میزان عمومیت داشتن یک قانون را با استفاده از رابطه (۳-۱۵) نشان می‌دهد.

تفسیر قوانین به تفصیل در بخش ۴-۵ تفسیر قوانین و نتایج حاصل از تحقیق ذکر گردیده است.

جدول شماره (۴-۱۳)- قوانین استخراجی و فاکتورهای پشتیبان، قطعیت و پوشش

	Rules	LHS Support	RHS Support	RHS Accuracy	RHS Coverage
1	%Delta Inflation Rate(1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	60	29,31	0.483333, 0.516667	1.0, 1.0
2	%Delta Crude Oil (1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	36	22,14	0.611111, 0.388889	0.758621, 0.451613
3	%Delta Crude Oil (-1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	24	7,17	0.291667, 0.708333	0.241379, 0.548387
4	%Delta Gold Price(-1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	29	14,15	0.482759, 0.517241	0.482759, 0.483871
5	%Delta Gold Price(1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	31	15,16	0.483871, 0.516129	0.517241, 0.516129
6	%Delta Currency Rate(1) => %Delta Saipa Stock Return(1) OR Delta Saipa Stock Return(-1)	37	17,20	0.459459, 0.540541	0.586207, 0.645161
7	%Delta Currency Rate(-1) => %Delta Saipa Stock Return(1) OR Delta Saipa Stock Return(-1)	23	12,11	0.521739, 0.478261	0.413793, 0.354839
8	%Delta coin price(-1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	19	11,8	0.578947, 0.421053	0.37931, 0.258065
9	%Delta coin price(1) => %Delta Saipa Stock Return(-1) OR %Delta Saipa Stock Return(1)	41	23,18	0.560976, 0.439024	0.741935, 0.62069
10	%Delta TEPIX(-1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	15	9,6	0.6, 0.4	0.741935, 0.62069
11	%Delta TEPIX(1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	45	20,25	0.444444, 0.555556	0.689655, 0.806452
12	%Delta TEDPIX(1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	31	20,11	0.645161, 0.354839	0.689655, 0.354839
13	%Delta TEDPIX(-1) => %Delta Saipa Stock Return(-1) OR %Delta Saipa Stock Return(1)	29	20,9	0.689655, 0.310345	0.645161, 0.310345
14	%Delta shakhes sanat(-1) => %Delta Saipa Stock Return(1) OR Delta Saipa Stock Return(-1)	29	12,17	0.413793, 0.586207	0.413793, 0.548387
15	%Delta shakhes sanat(1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	31	17,14	0.548387, 0.451613	0.586207, 0.451613
16	%Delta Paye Pooli(-1) => %Delta Saipa Stock Return(1) OR %Delta Saipa Stock Return(-1)	17	9,8	0.529412, 0.470588	0.310345, 0.258065
17	%Delta Paye Pooli(1) => %Delta Saipa Stock Return(-1) OR %Delta Saipa Stock Return(1)	43	23,20	0.534884, 0.465116	0.741935, 0.689655
18	%Delta shakhese maskan(1) => Delta Saipa Stock Return(1) OR Delta Saipa Stock Return(-1)	47	26,21	0.553191, 0.446809	0.896552, 0.677419
19	%Delta shakhese maskan(-1) => Delta Saipa Stock Return(1) OR Delta Saipa Stock Return(-1)	13	3,10	0.230769, 0.769231	0.103448, 0.322581

جدول شماره (۴-۱۴)- قوانین استخراجی و فاکتورهای پشتیبان، قطعیت و پوشش

	Rules	LHS Support	RHS Support	RHS Accuracy	RHS Coverage
20	% Delta TEPIX([0.70907, *)) => Return oscillation(-1) OR Return oscillation(1)	43	25,18	0.581395, 0.418605	0.806452, 0.62069
21	% Delta Inflation Rate([1.06826, *)) => Return oscillation(-1) OR Return oscillation(1)	42	25,17	0.595238, 0.404762	0.806452, 0.586207
22	% Delta shakhe maskan([0.45559, *)) => Return oscillation(-1) OR Return oscillation(1)	38	16,22	0.421053, 0.578947	0.516129, 0.758621
23	Delta TEDPIX([-1.02369, *)) => Return oscillation(1) OR Return oscillation(-1)	36	24,12	0.666667, 0.333333	0.827586, 0.387097
24	Delta TEDPIX([*, -1.02369)) => Return oscillation(-1) OR Return oscillation(1)	24	19,5	0.791667, 0.208333	0.612903, 0.172414
25	% Delta shakhe maskan([*, 0.45559)) => Return oscillation(-1) OR Return oscillation(1)	22	15,7	0.681818, 0.318182	0.483871, 0.241379
26	% Delta Inflation Rate([*, 1.06826)) => Return oscillation(1) OR Return oscillation(-1)	18	12,6	0.666667, 0.333333	0.413793, 0.193548
27	% Delta TEPIX([*, 0.70907)) => Return oscillation(-1) OR Return oscillation(1)	17	6,11	0.352941, 0.647059	0.193548, 0.37931
28	Delta TEDPIX([*, -3.40683)) => Return oscillation(-1) OR Return oscillation(1)	16	13,3	0.8125, 0.1875	0.419355, 0.103448
29	% Delta shakhe maskan([*, -0.10167)) => Return oscillation(-1) OR Return oscillation(1)	13	10,3	0.769231, 0.230769	0.322581, 0.103448
30	% Delta Crude Oil([*, -5.33526)) => Return oscillation(-1) OR Return oscillation(1)	9	6,3	0.666667, 0.333333	0.193548, 0.103448
31	% Delta shakhes sanat([*, -5.71580)) => Return oscillation(-1) OR Return oscillation(1)	9	6,3	0.666667, 0.333333	0.193548, 0.103448
32	% Delta Paye Pooli([*, -1.78992)) => Return oscillation(-1) OR Return oscillation(1)	8	2,6	0.25, 0.75	0.064516, 0.206897
33	Delta TEDPIX([*, -3.82799)) => Return oscillation(-1) OR Return oscillation(1)	15	13,2	0.866667, 0.133333	0.419355, 0.068966
34	% Delta shakhe maskan([*, 0.11647)) => Return oscillation(-1) OR Return oscillation(1)	15	11,4	0.733333, 0.266667	0.354839, 0.137931
35	% Delta Inflation Rate([*, 0.98840)) => Return oscillation(1) OR Return oscillation(-1)	15	11,4	0.733333, 0.266667	0.37931, 0.129032
36	% Delta Coin Price([5.30034, *)) => Return oscillation(1) OR Return oscillation(-1)	15	5,10	0.333333, 0.666667	0.172414, 0.322581
37	% Delta Coin Price([-1.30286, 1.88848)) => Return oscillation(-1) OR Return oscillation(1)	15	10,5	0.666667, 0.333333	0.322581, 0.172414
38	% Delta Coin Price([1.88848, 5.30034)) => Return oscillation(1) OR Return oscillation(-1)	15	10,5	0.666667, 0.333333	0.344828, 0.16129

جدول شماره (۴-۱۵)- قوانین استخراجی و فاکتورهای پشتیبان، قطعیت و پوشش

	Rules	LHS Support	RHS Support	RHS Accuracy	RHS Coverage
39	% Delta Inflation Rate([1.06826, *)) AND % Delta Crude Oil([*, 0.69723)) AND % Delta TEPIX([0.70907, *)) => Return oscillation(-1)	12	12	1.0	0.387097
40	Delta TEDPIX([-1.02369, *)) AND % Delta Paye Pooli([*, 1.11968)) AND % Delta shakhe maskan([0.45559, *)) => Return oscillation(1)	11	11	1.0	0.37931
41	% Delta Crude Oil([*, 0.69723)) AND % Delta TEPIX([0.70907, *)) AND Delta TEDPIX([*, -1.02369)) => Return oscillation(-1)	9	9	1.0	0.290323
42	% Delta Inflation Rate([1.06826, *)) AND % Delta TEPIX([0.70907, *)) AND Delta TEDPIX([*, -1.02369)) AND % Delta Paye Pooli([1.11968, *)) => Return oscillation(-1)	9	9	1.0	0.290323
43	% Delta Inflation Rate([1.06826, *)) AND % Delta TEPIX([0.70907, *)) AND Delta TEDPIX([*, -1.02369)) AND % Delta shakhe maskan([0.45559, *)) => Return oscillation(-1)	9	9	1.0	0.290323
44	% Delta Crude Oil([0.69723, *)) AND % Delta TEPIX([0.70907, *)) AND Delta TEDPIX([-1.02369, *)) AND % Delta Paye Pooli([*, 1.11968)) => Return oscillation(1)	9	9	1.0	0.310345
45	% Delta Gold Price([-0.32577, 1.57707)) => Return oscillation(1)	7	7	1.0	0.241379
46	% Delta Paye Pooli([2.78111, 4.44954)) => Return oscillation(-1)	7	7	1.0	0.225806
47	% Delta Coin Price([-5.50813, -2.34889)) => Return oscillation(1)	7	7	1.0	0.241379
48	% Delta Gold Price([5.68560, 9.79819)) => Return oscillation(-1)	7	7	1.0	0.225806
49	% Delta Inflation Rate([0.98840, 1.35166)) AND % Delta Crude Oil([*, -4.39659)) => Return oscillation(-1)	6	6	1.0	0.193548
50	% Delta Coin Price([1.88848, 5.30034)) AND % Delta shakhe maskan([0.11647, 0.93419)) => Return oscillation(1)	6	6	1.0	0.206897
51	Delta TEDPIX([*, -3.82799)) AND % Delta shakhe maskan([0.93419, 1.80299)) => Return oscillation(-1)	6	6	1.0	0.193548

جدول شماره (۴-۱۶)- قوانین استخراجی و فاکتورهای پشتیبان، قطعیت و پوشش

	Rules	LHS Support	RHS Support	RHS Accuracy	RHS Coverage
52	Delta TEPIX(1) AND Delta shakhes maskan (-1) => liquidity(1)	10	10	1.0	0.27027
53	Delta shakaskanhes m(-1) AND Delta Saipa Stock Return(-1) => liquidity(1)	10	10	1.0	0.27027
54	Delta Paye Pooli(-1) AND Delta Saipa Stock Return(-1) => liquidity(1)	8	8	1.0	0.216216
55	Delta Crude Oil (1) AND Delta shakaskanhes m(-1) => liquidity(1)	8	8	1.0	0.216216
56	Delta Paye Pooli(1) AND Delta shakaskanhes m(-1) => liquidity(1)	8	8	1.0	0.216216
57	% Delta Crude Oil([*, -4.20942)) AND % Delta TEPIX([*, 2.09897)) => liquidity(1)	6	6	1.0	0.162162
58	% Delta shakhe maskan([*, 0.70157)) AND Delta Saipa Stock Return([*, -8.37292)) => liquidity(1)	8	8	1.0	0.216262
59	% Delta Coin Price([1.98690, 2.95438)) => liquidity(1)	8	8	1.0	0.216262
60	Delta Saipa Stock Return([1.00101, 7.17423)) => liquidity(1)	8	8	1.0	0.216262
61	Delta Saipa Stock Return([1.00101, 7.17423)) => liquidity(1)	8	8	1.0	0.216216
62	% Delta Coin Price([1.98690, 2.95438)) => liquidity(1)	8	8	1.0	0.216216
63	% Delta Currency Rate([2.35680, *)) AND % Delta shakhe maskan([*, 0.11647)) => liquidity(1)	7	7	1.0	0.189189
64	% Delta Gold Price([*, -2.52742)) AND % Delta shakhe maskan([*, 0.11647)) => liquidity(1)	7	7	1.0	0.189189
65	% Delta shakhe maskan([*, 0.11647)) AND Delta Saipa Stock Return([-10.53840, -0.17454)) => liquidity(1)	7	7	1.0	0.189189
66	Delta Saipa Stock Return([-0.17454, 10.30530)) => liquidity(3) OR liquidity(2) OR liquidity(1)	15	3,3,9	0.6	0.43
67	Delta Saipa Stock Return([*, -10.53840)) => liquidity(2) OR liquidity(1) OR liquidity(4) OR liquidity(3)	15	3,10,1,1	0.2	0.5
68	Delta Saipa Stock Return([10.30530, *)) => liquidity(1) OR liquidity(2) OR liquidity(5) OR liquidity(4) OR liquidity(3)	15	8,4,1,1,1	0.53	0.5
69	Delta Saipa Stock Return([-10.53840, -0.17454)) => liquidity(3) OR liquidity(2) OR liquidity(1) OR liquidity(5)	15	2,2,10,1	0.67	0.5
70	Delta TEDPIX([*, -3.82799)) => liquidity(2) OR liquidity(1) OR liquidity(3) OR liquidity(5) OR liquidity(4)	15	5,7,1,1,1	0.47	0.42

۴-۴-۲-۴ ارزیابی مدل با تکنیک K-Fold Cross Validation

سنجش نحوه‌ی عملکرد مدل‌های پیش‌بینی معمولاً بدین صورت انجام می‌شود که کاربر بخشی از پایگاه داده را به عنوان مجموعه آموزشی[1] و بخشی دیگر را به عنوان مجموعه آزمایشی[2] در نظر می‌گیرد و با استفاده از الگوریتم‌های مختلف قوانین حاکم بر مجموعه آموزشی را استخراج نموده و آنان را به منظور ارزیابی مدل در مجموعه تست، آزمایش می‌نماید و بدین ترتیب میزان دقت مدل پیش‌بینی تخمین زده می‌شود.

یکی از معایب موجود در روش فوق از دست دادن بخشی از قوانین حاکم بر پایگاه داده می‌باشد، زیرا مجموعه‌ای که به عنوان مجموعه تست در نظر گرفته می‌شود، خود دارای قوانینی است که نادیده گرفته می‌شود. به منظور برطرف کردن این نقص از تکنیک‌های ارزیابی مدل دیگری همچون Holdout, Bootstrap, Random Subsampling, و K-Fold Cross Validation استفاده می‌شود. که در این پژوهش از روش ارزیابی مقطعی (Cross Validation) یا اعتبارسنجی ضربدری استفاده شده است.

اعتبارسنجی ضربدری، که گاهی تخمین گردشی نیز نامیده می‌شود، یک روش ارزیابی است که مشخص می‌کند نتایج یک تحلیل آماری بر روی یک مجموعه داده تا چه اندازه قابل تعمیم و مستقل از داده‌های آموزشی است. این تکنیک بطور ویژه در کاربردهای پیش بینی مورد استفاده قرار می‌گیرد تا مشخص شود مدل موردنظر تا چه اندازه در عمل مفید خواهد بود. بطور کلی یک دور از اعتبارسنجی ضربدری شامل افراز داده‌ها به دو زیرمجموعه مکمل، انجام تحلیل بر روی یکی از آن زیرمجموعه‌ها (داده‌های آموزشی) و اعتبارسنجی تحلیل با استفاده از داده‌های مجموعه دیگر است (داده‌های اعتبارسنجی یا تست). برای کاهش پراکندگی، عمل اعتبارسنجی چندین بار با افرازهای مختلف انجام و از نتایج اعتبارسنجی ها میانگین گرفته می شود.

[1] : Training Set
[2] : Test Set

در روش K-Fold نوع اعتبارسنجی داده‌ها به K زیرمجموعه افراز می‌شوند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و K-1 تای دیگر برای آموزش بکار می‌روند. این روال K بار تکرار می‌شود و همه داده‌ها دقیقا یک‌بار برای آموزش و یک‌بار برای اعتبارسنجی بکار می‌روند. در هر بار اعتبارسنجی نتایج تحت عنوان جدولی نشان داده می‌شود و در نهایت میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده می‌شود. البته می‌توان از روش‌های دیگر برای ترکیب نتایج استفاده کرد. بطور معمول از 10- Fold استفاده می‌شود[۹۳,۹۴,۹۵].

نرم‌افزار ROSETTA برای اجرای تکنیک ارزیابی مقطعی از یک Meta-Algorithm استفاده می‌کند و به کاربر اجازه می‌دهد که الگوریتم‌های مختلفی در مراحل مختلف کار استفاده نماید.

با اجرای الگوریتم Cross Validation و ارزیابی مدل میانگین دقت تخمین مدل پیش‌بینی برای الگوریتم ژنتیک ۰/۸۸۳۳ و برای الگوریتم هولت ۰/۷۳۳۳ می‌باشد.

۴-۵ تفسیر قوانین و نتایج حاصل از تحقیق

۱- شاخص بهای کالاها و خدمات مصرفی:

این شاخص در کل دوره‌ی مورد بررسی تنها افزایش یافته و روند کاهشی را تجربه نکرده، با توجه به این شرایط روند داده‌های ورودی هر ۶۰ ماه افزایشی بوده که در ۰/۴۸ درصد موارد (۲۹ مورد) با افزایش بازده سهام و ۰/۵۲ درصد موارد (۳۱ مورد) با کاهش بازده سهام همراه بوده است، بنابراین می‌توان چنین نتیجه گرفت که شاخص بهای کالاها و خدمات مصرفی تأثیر مهمی در روند بازده سهام ندارد.

رفتار بازده سهام در مورد شاخص بهای کالاها و خدمات مصرفی تا حد زیادی مطابق با رفتار شاخص قیمت و بازده نقدی (TEDPIX) در مورد این شاخص می‌باشد، در حالی که شاخص قیمت بورس اوراق بهادار (TEPIX) رفتاری کاملاً متفاوت از خود نشان می‌دهد و در ۷۵ درصد موارد با افزایش شاخص بهای کالاها و خدمات مصرفی، افزایش یافته است.

۲- قیمت جهانی نفت خام:

این شاخص در کل دوره‌ی مورد بررسی در ۳۶ مورد افزایش و در ۲۴ مورد کاهش داشته است.

در زمانی که با افزایش همراه بوده است در ۶۱ درصد موارد (۲۲ مورد از ۳۶ مورد) موجب افزایش بازده و در ۳۹ درصد موارد (۱۴ مورد از ۳۶ مورد) با کاهش بازده همراه بوده است. با توجه به اختلاف کم میان موارد افزایش و کاهش بازده (۶۱ درصد در مقابل ۳۹ درصد)، می‌توان گفت که افزایش قیمت نفت خام *تقریباً* منجر به افزایش بازده سهام می‌شود.

در زمانی که قیمت جهانی نفت خام روند کاهشی را تجربه کرده است، در ۷۱ درصد موارد (۱۷ مورد از ۲۴ مورد) بازده سهام نیز کاهش یافته و تنها در ۲۹ درصد موارد (۷ مورد از ۲۴ مورد) بازده سهام افزایش یافته است، بنابراین می‌توان نتیجه گرفت که کاهش قیمت نفت خام می‌تواند منجر به کاهش بازده سهام گردد.

بازده سهام سایپا در این مورد بیشتر از شاخص قیمت و بازده نقدی (TEDPIX) تبعیت می‌کند تا از شاخص کل بورس اوراق بهادار (TEPIX)، زیرا TEDPIX روند افزایشی و کاهشی مطابق با بازده سهام دارد در حالی که TEPIX با کاهش قیمت نفت خام تنها در ۲۱ درصد موارد کاهش یافته و در ۷۹ درصد موارد با افزایش همراه بوده است.

۳- قیمت جهانی طلا

قیمت جهانی طلا در کل دوره‌ی مورد بررسی در ۳۱ مورد افزایش و در ۲۹ مورد کاهش داشته است.

در زمانی که با افزایش همراه بوده است در ۴۸ درصد موارد (۱۵ مورد از ۳۱ مورد) افزایش بازده سهام دیده شده و در ۵۲ درصد موارد (۱۶ مورد از ۳۱ مورد) کاهش بازده سهام مشاهده شده است، بنابراین افزایش قیمت طلا تأثیری در افزایش یا کاهش بازده سهام سایپا ندارد.

هنگامی که قیمت جهانی طلا کاهش یافته است، در ۴۸ درصد موارد (۱۴ مورد از ۲۹ مورد) بازده سهام افزایش یافته و در ۵۲ درصد موارد (۱۵ مورد از ۲۹ مورد) کاهش یافته است. بنابراین کاهش قیمت طلا مانند افزایش آن تأثیری در تغییرات بازده سهام شرکت سایپا ندارد.

رفتار شاخص قیمت و بازده نقدی (TEDPIX) در این مورد کاملاً شبیه به رفتار بازده سهام شرکت سایپا است در حالی که شاخص کل قیمت بورس (TEPIX) نه تنها با افزایش قیمت جهانی طلا در ۷۴ درصد موارد افزایش می‌یابد بلکه با کاهش آن نیز در ۷۶ درصد موارد افزایش یافته است و به نظر می‌رسد که رفتاری مستقل از نوسان قیمت جهانی طلا، اما متفاوت با شاخص قیمت و بازده نقدی از خود نشان می‌دهد.

۴- متوسط قیمت فروش دلار در بازار آزاد تهران

متوسط نرخ فروش دلار در بازار تهران در کل دوره‌ی مورد بررسی در ۳۷ مورد افزایش و در ۲۳ مورد کاهش داشته است.

با افزایش نرخ فروش دلار، بازده سهام سایپا در ۵۴ درصد موارد (۲۰ مورد از ۳۷ مورد) افزایش یافته و در ۴۶ درصد موارد (۱۷ مورد از ۳۷ مورد) کاهش یافته است، بنابراین افزایش نرخ فروش دلار تأثیری در تغییرات بازده سهام ندارد.

هنگامی که نرخ فروش دلار کاهش یافته است، بازده سهام سایپا در ۴۸ درصد موارد (۱۱ مورد از ۲۳ مورد) کاهش یافته و در ۵۲ درصد از موارد (۱۲ مورد از ۲۳ مورد) افزایش یافته است. بنابراین کاهش این نرخ همانند افزایش آن تأثیری در تغییرات بازده سهام سایپا ندارد.

رفتار شاخص قیمت و بازده نقدی (TEDPIX) در این مورد کاملاً شبیه به رفتار بازده سهام شرکت سایپا است در حالی که شاخص کل قیمت بورس (TEPIX) نه تنها با افزایش نرخ فروش دلار در ۶۸ درصد موارد افزایش می‌یابد بلکه با کاهش آن نیز در ۸۷ درصد موارد افزایش یافته است و به نظر می‌رسد که رفتاری مستقل از نوسان نرخ فروش دلار در بازار آزاد تهران، اما متفاوت با شاخص قیمت و بازده نقدی از خود نشان می‌دهد.

۵- متوسط قیمت فروش یک سکه تمام بهار آزادی (طرح قدیم) در بازار آزاد شهر تهران

متوسط قیمت فروش یک سکه در بازار تهران در کل دوره‌ی مورد بررسی در ۴۱ مورد افزایش و در ۱۹ مورد کاهش داشته است.

هنگامی که قیمت سکه افزایش می‌یابد، بازده سهام در ۵۶ درصد موارد (۲۳ مورد از ۴۱ مورد) کاهش یافته و در ۴۴ درصد موارد (۱۸ مورد از ۴۱ مورد) افزایش یافته است.

از طرفی با کاهش قیمت سکه، بازده سهام در ۵۸ درصد موارد (۱۱ مورد از ۱۹ مورد) افزایش یافته و در ۴۲ درصد موارد (۸مورد از ۱۹ مورد) کاهش یافته است. بنابراین می‌توان نتیجه گرفت که نوسانات بازده

سهام سایپا ارتباط معکوس ضعیفی با قیمت سکه دارد.

شاخص قیمت و بازده نقدی دقیقاً همانند بازده سهام سایپا ارتباط معکوس ضعیفی با قیمت سکه دارد به نحوی که با کاهش قیمت سکه در ۶۳ درصد موارد افزایش و با افزایش قیمت سکه در ۵۴ درصد موارد کاهش یافته است. در حالی که شاخص کل قیمت با افزایش یا کاهش قیمت سکه به ترتیب در ۷۱ درصد موارد افزایش و در ۸۴ درصد موارد افزایش یافته است.

۶- پایه پولی

این شاخص در کل دوره‌ی مورد بررسی در ۴۳ مورد افزایش و در ۱۷ مورد کاهش داشته است.

با افزایش این شاخص بازده سهام به ترتیب در ۵۴ درصد از موارد (۲۳ مورد از ۴۳ مورد) و ۴۶ درصد از موارد (۲۰ مورد از ۴۳ مورد) کاهش و افزایش یافته است.

با کاهش این شاخص بازده سهام سایپا در ۵۳ درصد از موارد (۹ مورد از ۱۷ مورد) افزایش یافته و در ۴۷ درصد از موارد (۸ مورد از ۱۷ مورد) کاهش یافته است. بنابراین نوسانات این شاخص تأثیری در تغییرات بازده سهام ندارد.

رفتار شاخص قیمت و بازده نقدی بورس در زمینه شاخص پایه پولی دقیقاً مطابق با رفتار بازده سهام سایپا بوده است در حالی که شاخص کل قیمت با افزایش یا کاهش این شاخص افزایش یافته است.

۷- شاخص مسکن

این شاخص در کل دوره‌ی مورد بررسی در ۴۷ مورد افزایش و در ۱۳ مورد کاهش داشته است.

با افزایش شاخص مسکن، بازده سهام شرکت سایپا در ۵۵ درصد موارد (۲۶ مورد از ۴۷ مورد) افزایش و در ۴۶ درصد موارد (۲۱ مورد از ۴۷ مورد) کاهش یافته است و با کاهش این شاخص در ۲۳ درصد از موارد (۳ مورد از ۱۳ مورد) افزایش و در ۷۷ درصد از موارد (۱۰ مورد از ۱۳ مورد) کاهش یافته است. بنابراین می‌توان گفت که رفتار بازده سهام با کاهش شاخص مسکن کاهش می‌یابد و با افزایش شاخص

مسکن تقریباً افزایش می‌یابد.

شاخص قیمت و بازده نقدی با کاهش این شاخص کاهش می‌یابد ولی با افزایش آن، همچون بازده سهام سایپا عمل نکرده و در ۵۳ درصد از موارد کاهش می‌یابد. البته به صورت کلی تفاوت رفتاری میان شاخص قیمت و بازده نقدی و بازده سهام سایپا به دلیل کم بودن قابل چشم‌پوشی می‌باشد.

شاخص کل قیمت بدون توجه به نوسانات این شاخص، چه در حالت افزایش و چه در حالت کاهش شاخص مسکن، افزایش می‌یابد.

۸- شاخص قیمت بورس اوراق بهادار تهران (TEPIX)

این شاخص در طی دوره مورد بررسی در ۴۵ مورد افزایش و در ۱۵ مورد کاهش داشته است.

با افزایش TEPIX بازده سهام در ۶۰ درصد موارد (۲۵ مورد از ۴۵ مورد) کاهش یافته است و در ۴۰ درصد موارد (۲۰ مورد از ۴۰ مورد) افزایش یافته است و با کاهش این شاخص در ۶۰ درصد موارد (۹ مورد از ۱۵ مورد) بازده سهام افزایش یافته و در ۴۰ درصد موارد (۶ مورد از ۱۵ مورد) کاهش یافته است.

رفتار شاخص قیمت و بازده نقدی در مورد TEPIX کاملاً با رفتار بازده سهام سایپا مطابق است یعنی با افزایش TEPIX در ۵۳ درصد موارد کاهش داشته و با کاهش TEPIX در ۴۷ درصد موارد افزایش یافته است.

۹- شاخص قیمت و بازده نقدی بورس اوراق بهادار تهران (TEDPIX)

این شاخص در طول کل دوره‌ی مورد بررسی در ۳۱ مورد افزایش و در ۲۹ مورد کاهش داشته است.

با کاهش TEDPIX بازده سهام شرکت سایپا در ۶۹ درصد موارد (۲۰ مورد از ۲۹ مورد) کاهش داشته و در ۳۱ درصد موارد (۹ مورد از ۲۹ مورد) افزایش داشته است. با افزایش TEDPIX بازده سهام شرکت سایپا در ۶۵ درصد موارد (۲۰ مورد از ۳۱ مورد) افزایش یافته است و در ۳۵ درصد موارد (۱۱ مورد از ۲۹ مورد) کاهش یافته است. بنابراین می‌توان گفت که بازده سهام شرکت سایپا همگام با شاخص قیمت و

بازده نقدی حرکت می‌کند.

رفتار شاخص قیمت بورس اوراق بهادار تهران همچنان مستقل از TEDPIX با افزایش یا کاهش TEDPIX افزایش می‌یابد.

۱۰- شاخص صنعت خودروسازی

این شاخص در طول کل دوره‌ی مورد بررسی در ۳۱ مورد افزایش و در ۲۹ مورد کاهش داشته است با افزایش شاخص صنعت، بازده سهام سایپا در ۵۵ درصد موارد (۱۷ مورد از ۳۱ مورد) افزایش و در ۴۵ درصد موارد (۱۴ مورد از ۳۱ مورد) کاهش یافته است.

با کاهش شاخص صنعت، بازده سهام سایپا در ۵۹ درصد موارد (۱۷ مورد از ۲۹ مورد) کاهش و در ۴۱ درصد موارد (۱۲مورد از ۲۹ مورد) افزایش یافته است. بنابراین مشخص می‌شود که بازده سهام سایپا چندان از شاخص صنعت خودروسازی تبعیت نمی‌کند.

۱۱- نقدشوندگی

نقدشوندگی سهام شرکت سایپا بیشترین ارتباط و همبستگی را با عامل شاخص قیمت مسکن دارد به نحوی که در ۸۲ درصد از مواردی که نقدشوندگی در سطح متغیر کلامی خیلی خوب قرار دارد، شاخص قیمت مسکن روندی نزولی را سپری می‌کرده است. این موضوع بیان‌گر سرمایه‌گذاری وجوه سرگردان دربازار بورس در زمانی است که رونق بازار مسکن و ... کاهش یافته است. علاوه بر آن در قوانین مشاهده شده است که نقدشوندگی سهام با کاهش بازده سهام افزایش می‌یابد.

فصل پنجم

نتیجه گیری

۵-۱ مقدمه

اهیمت بازده سهام در بورس اوراق بهادار باعث شده که موضوع تعداد زیادی از تحقیقات مالی را به خود اختصاص دهد. بخش بسیاری از این تحقیقات به پیش‌بینی بازده سهام با استفاده از تکنیک‌ها و شاخص‌های متفاوت پرداخته‌اند. در این پژوهش رفتار بازده سهام با استفاده از تکنیک مجموعه راف مورد ارزیابی قرار گرفته و عوامل مؤثر در پیش‌بینی آن شناسایی شده است. در این بخش محدودیت‌های موجود در انجام پژوهش، نتایج کلی و پیشنهادات ارائه می‌گردد.

۵-۲ نتیجه‌گیری

۱. تغییرات شاخص بهای کالاها و خدمات مصرفی (CPI) در بازده سهام شرکت سایپا مؤثر نمی‌باشد، بنابراین می‌توان در تخمین پیش‌بینی‌های بازده این سهام از این شاخص صرف نظر کرد. نتایج به‌دست آمده از قوانین استخراجی در مورد شاخص قیمت و بازده نقدی با بازده سهام سایپا مطابقت دارد، در حالی که روند شاخص کل قیمت بورس اوراق بهادار کاملاً متفاوت بوده است.

یکی از دلایل بی‌تأثیر بودن نرخ بهای کالاها و خدمات مصرفی می‌تواند عدم تعدیل مناسب سود شرکت سایپا با توجه به روند افزایش قیمت‌ها باشد. زیرا این شرکت در قیمت‌گذاری محصولات خود تماماً مختار نبوده و افزایش قیمت‌ها به صورت کنترل‌شده اعمال می‌گردد، بنابراین قیمت سهام این شرکت با توجه به افزایش نرخ بهای کالاها و خدمات مصرفی افزایش نمی‌یابد.

۲. افزایش قیمت جهانی نفت خام، تقریباً با افزایش بازده سهام همراه بوده و کاهش آن منجر به کاهش بازده سهام سایپا می‌گردد. بازده سهام سایپا در این مورد از شاخص قیمت و بازده نقدی (TEDPIX) تبعیت می‌کند.

با توجه به این‌که وضعیت اقتصادی ایران بسیار به نفت وابسته می‌باشد، افزایش و کاهش بازده سهام شرکت سایپا با تغییرات قیمت جهانی نفت کاملاً قابل توجیه است. با افزایش قیمت جهانی

نفت و بهبود وضعیت آن، شرایط اقتصادی بهتر شده و تولید شرکت افزایش می‌یابد.

۳. تغییرات قیمت جهانی طلا تأثیری در افزایش یا کاهش بازده سهام شرکت سایپا ندارد. رفتار شاخص قیمت و بازده نقدی دقیقاً مشابه با رفتار بازده سهام شرکت سایپا بوده ولی رفتار شاخص کل قیمت متفاوت از این دو بوده است.

۴. تغییرات متوسط قیمت فروش دلار در بازار آزاد تهران در افزایش یا کاهش بازده سهام شرکت سایپا بی‌تأثیر می‌باشد. شاخص قیمت و بازده نقدی همچون بازده سهام سایپا مستقل از نوسانات این شاخص است. شاخص کل قیمت علاوه بر مستقل بودن از این شاخص، تفاوت رفتاری زیادی با شاخص قیمت و بازده نقدی و بازده سهام سایپا دارد.

۵. نوسانات بازده سهام سایپا ارتباط معکوس ضعیفی با متوسط قیمت فروش یک سکه بهار آزادی (طرح قدیم) در بازار تهران دارد. رفتار شاخص قیمت و بازده نقدی در زمینه‌ی این شاخص مطابق با رفتار بازده سهام سایپا است در حالی که رفتار شاخص کل قیمت مستقل از این شاخص می‌باشد. در حالت عمومی رفتار طلا در جهتی خلاف جهت حرکت عمومی اقتصاد است و به همین علت یک Hedge Asset شناخته می‌شود.

۶. نوسانات شاخص پایه پولی هیچ تأثیری در افزایش یا کاهش بازده سهام سایپا ندارد و می‌توان از این شاخص در پیش‌بینی‌های آتی صرف نظر کرد. رفتار شاخص قیمت و بازده نقدی در مورد این شاخص دقیقاً مطابق با رفتار بازده سهام سایپا بوده در حالی که شاخص کل قیمت رفتاری متفاوت از خود نشان داده است.

۷. بازده سهام سایپا با افزایش شاخص مسکن تقریباً افزایش یافته و با کاهش آن، تقلیل می‌یابد. رفتار شاخص قیمت و بازده نقدی تا حد زیادی به رفتار بازده سهام نزدیک بوده است در حالی که نحوه‌ی عملکرد شاخص کل قیمت متفاوت می‌باشد.

افزایش شاخص مسکن دلالت دارد به بهبود شرایط کلی اقتصادی کشور، بنابراین با افزایش درآمدها بازار مسکن، خودرو، لوازم منزل و ... رونق یافته و بازار بورس نیز با اقبال مواجه می‌شود.

۸. بازده سهام سایپا رابطه‌ی معکوس ضعیفی با شاخص قیمت بورس اوراق بهادار (TEPIX) دارد. رفتار شاخص قیمت و بازده نقدی در مورد این شاخص بسیار مشابه با رفتار بازده سهام سایپا می‌باشد.

۹. بازده سهام شرکت سایپا با افزایش شاخص قیمت و بازده نقدی بورس اوراق بهادار تهران (TEDPIX) افزایش و با کاهش آن کاهش می‌یابد. بنابراین می‌توان گفت که این شاخص در پیش‌بینی‌های آینده می‌تواند ملاک مناسبی قرار گیرد.

۱۰. بازده سهام شرکت سایپا با افزایش شاخص صنعت خودروسازی افزایش می‌یابد و با کاهش آن کاهش می‌یابد. اما به دلیل ارتباط ضعیفی که بین این دو شاخص وجود دارد نمی‌توان آن را مبنایی برای پیش‌بینی‌های آتی قرار داد.

۱۱. نقدشوندگی سهام شرکت سایپا بیشترین ارتباط و همبستگی را با عامل شاخص قیمت مسکن دارد به نحوی که در ۸۲ درصد از مواردی که نقدشوندگی در سطح متغیر کلامی خیلی خوب قرار دارد، شاخص قیمت مسکن روندی نزولی را سپری می‌کرده است. با توجه به اینکه شاخص قیمت مسکن نمادی از رونق بازار است، چنین به نظر می‌رسد که با کاهش رونق بازار گرایش سرمایه‌گذاران به سرمایه‌گذاری وجوه سرگردان در بازار بورس افزایش می‌یابد، بنابراین تعداد معاملات افزایش یافته و نقدشوندگی سهام بیشتر می‌شود.

نتایج حاصل از این پژوهش حاکی از آن است که بازده سهام سایپا بیش از هر عاملی از شاخص قیمت و بازده نقدی بورس اوراق بهادار تهران تبعیت می‌کند و علاوه بر آن قدرت و توانایی تکنیک راف ست در استخراج قوانین و رفتار پنهان حاکم بر سهام مورد بررسی، ارزیابی گردید.

۵-۳ محدودیت‌های تحقیق

۱. کمبود اطلاعات علمی برای دست‌یابی به مفاهیم اولیه مجموعه راف در مطالعات داخلی
۲. بخشی از محدودیت‌های موجود در این پژوهش ناشی از کمبود اطلاعات در زمینه نحوه استفاده از نرم‌افزار ROSETTA می‌باشد.
۳. علاوه بر آن محدودیت جمع‌آوری برخی از داده‌های اقتصادی به صورت ماهانه نیز وجود دارد.

۵-۴ پیشنهادات

۵-۴-۱ پیشنهادات کاربردی

۱. سرمایه‌گذاران برای ساختن پرتفلیو خود می‌توانند رفتار سهام مورد نظر را با استفاده از راف ست مورد بررسی قرار دهند و در جهت کاهش ریسک سرمایه‌گذاری و افزایش میزان نقدشوندگی از نتایج آن بهره‌مند گردند.
۲. به سرمایه‌گذارانی که تمایل به معامله سهام سایپا دارند پیشنهاد می‌گردد که بیش از هر عامل دیگری به روند شاخص قیمت و بازده نقدی بورس اوراق بهادار تهران (TEDPIX) و نوسانات قیمت نفت خام به منظور پیش‌بینی بازده سهام توجه نمایند.

۵-۴-۲ پیشنهادات برای تحقیقات آتی

۱. استفاده از سایر شاخص‌های اقتصاد کلان امکان دارد که بر افزایش دقت تکنیک راف ست بیافزاید، بنابراین پیشنهاد می‌شود که در مطالعات آتی از سایر متغیرهای اقتصاد کلان استفاده گردد و رفتار بازده سهام با توجه به آنان بررسی شود.
۲. استفاده از متغیری که توانایی دخالت دادن شرایط سیاسی حاکم بر کشور را دارا باشد، می‌تواند روند کشف الگوهای پنهان رفتاری سهام شرکت‌های مختلف را تسهیل بخشد.

۳. استفاده از راف ست به همراه سایر تکنیک‌های محاسبات نرم از جمله سیستم‌های خاکستری، فازی، شبکه عصبی و ... می‌تواند نتایج بهتری را ارائه دهد.

۴. بررسی رفتار سایر سهام (از صنایع مختلف) و مقایسه‌ی آنان با یکدیگر می‌تواند روند رفتاری سهام صنایع مختلف را بررسی و الگوهای پنهان در آنان را کشف کند.

۵. استفاده از سایر معیارهای نقدشوندگی و مقایسه آن‌ها با یکدیگر و بررسی این‌که کدام معیار بیشترین توانایی را در نشان دادن سطح نقدشوندگی دارا می‌باشد.

منابع

منابع :

۱. جونز، چارلز پارکر (۱۳۸۶).مدیریت سرمایه‌گذاری. مترجمان تهرانی، رضا و عسگر نوربخش. چاپ سوم، انتشارات نگاه دانش، ص.۱۱

۲. دکتر رحمانی، علی و فرشته سعیدی(۱۳۸۷)؛ ارزیابی عملکرد مدل های لاجیت در پیش بینی بازده سهام، فصلنامه بورس اوراق بهادار، سال اول، شماره ۲، صفحه ۴۳-۸۵.

۳. زرا نژاد، منصور، مریم احمدی فرد و فخریه ساده نژاد(۱۳۹۱)؛ بکارگیری سیستم استنتاج فازی و شبکه های عصبی مصنوعی در پیش بینی بازده سهام، اولین همایش ملی کاربرد سیستم های هوشمند(محاسبات نرم) در علوم و صنایع.

۴. مهرانی، ساسان و کاوه مهرانی(۱۳۸۲). رابطه بین نسبت‌های سودآوری و بازده سهام در بازار بورس تهران، بررسی‌های حسابداری و حسابرسی، سال دهم، شماره ۳۳، ص ۹۳-۱۰۵.

۵. کرمی، غلامرضا و دیگران(۱۳۸۵). "بررسی روابط خطی و غیر خطی بین نسبت‌های مالی و بازده سهام در بورس اوراق بهادار تهران"، بررسی‌های حسابداری و حسابرسی، شماره ۴۶، ص۱۹-۴۶.

۶. ایزدی نیا، ناصر و منیژه رامشه و سعید یادگاری (۱۳۹۱)؛ پیش بینی جهت بازده سهام بر اساس حجم معاملات سهام؛ فصلنامه علمی پژوهشی حسابداری مالی؛ سال چهارم؛ شماره ۱۶؛ صفحه ۱۶۰-۱۷۴.

۷. اسلامی بیدگلی، غلامرضا و اعظم هنردوست(۱۳۹۱). مدل سه عاملی فاما و فرنچ و ریسک نقدشوندگی: شواهدی از بازار بورس اوراق بهادار تهران، فصلنامه دانش سرمایه گذاری، سال اول، شماره دوم.

۸. شریعت پناهی سید مجید، جعفری ابوالفضل(۱۳۸۹). مدیریت سرمایه‌گذاری، انتشارات اتحاد، چاپ دوم، صفحه ۳۹۸-۴۱۹.

۹. ثقفی، علی و محمد جواد سلیمی(۱۳۸۴)؛ متغیرهای بنیادی حسابداری و بازده سهام، مجله علوم اجتماعی و انسانی دانشگاه شیراز، دوره ی بیست و دوم، شماره دوم.

۱۰. ثقفی، علی و محمد جواد سلیمی(۱۳۸۴)؛ متغیرهای بنیادی حسابداری و بازده سهام، مجله علوم اجتماعی و انسانی دانشگاه شیراز، دوره ی بیست و دوم، شماره دوم.

۱۱. پور زمانی، زهرا و علی بشیری(۱۳۹۲)؛ آزمون مدل کارهارت برای پیش بینی بازده مورد انتظار به تفکیک سهام رشدی و ارزشی، مجله مهندسی مالی و مدیریت اوراق بهادار، شماره شانزدهم، پاییز ۱۳۹۲.

۱۲. نمازی، محمد و محمدمهدی کیامهر(۱۳۸۶). پیش‌بینی بازده روزانه سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از شبکه‌های عصبی مصنوعی، تحقیقات مالی، شماره ۲۴، صفحه ۱۱۵ تا ۱۳۴.

13. Wyss R. (2004) Measuring and Predicting Liquidity in the Stock Market. PHD Dissertation, University of St. Gallen.
14. Kang, W. & Zhang, H. (2014) Measuring liquidity in emerging markets. Pacific-Basin Finance Journal, 27, 49-71
15. Kim, S. & Lee, K. (2011) Pricing of liquidity risks by alternative liquidity measures. Confrence.
16. Fong, K., Holden, C.W. & Trzcinka, C. A. (2011) What are the best liquidity proxies for global research, Social Science Research Network. Social Science Electronic Publishing.
17. Lou, X. & Shu, T. (2014). Price impact or trading volume: why is the Amihud (2002) illiquidity measure priced, Social Science Research Network. Social Science Electronic Publishing.
18. Lesmond, David A. (2005). Liquidity of emerging markets, Journal of Financial Economics, 77, 411-452.

19. .Goyenko, R. Y., Holden, C. W. & Trzcinka, C. A. (2009). Do liquidity measures measure liquidity?. Journal of Financial Economics. 92, 153-181.
20. Pereira da Silva, P. (2013). Modeling and forecasting liquidity using principal component analysis and dynamic factor models. Working paper CMVM, CMVM-Portuguese securities commission.
21. Abderrazak, (.1122.) Day-of-the-week and the returns distribution: evidence from the Tunisian Stock Market, May 5-6
22. Akrami, H., Garkaz M. & Mehrazin, A., (2012). The effect of Ramadhan month on stocks abnormal return of the companies accepted in Tehran stock exchange. Economics and Finance Review (EFR), ISSN: 2047-0401.
23. Husain, F. (1998). A Seasonality in the Pakistan Equity Market: The Ramadhan Effect. Pakistan Development Review 7, 77-81.
24. Nilgun, C., Burak, G., & Burcu, K. (2008). The Month and Holy Days Effects on the Volatility of trade Deficit. Evidence from Turkey. Journal of Economic and Social Vol, 67-84.
25. Oguzsoy, C.B., Guven, B. (2004). Holy Days Effect on Istanbul Stock Exchange. Journal of Emerging Market Finance 3, 63-76.
26. Seyyed, F.J., Abraham, A. & Al-Hajji, M. (2005). Seasonality in Stock Returns and Volatility. Research in International Business and Finance 19: 374-383.
27. Robertson, R., & Combs, A. (Eds.). (2014). Chaos theory in psychology and the life sciences. Psychology Press.
28. Hilborn, R. C. (2004). Sea gulls, butterflies, and grasshoppers: A brief history of the butterfly effect in nonlinear dynamics. American Journal of Physics, 72(4), 425-427.
29. Xiu, D., & Karniadakis, G. E. (2002). The Wiener--Askey polynomial chaos for stochastic differential equations. SIAM Journal on Scientific Computing, 24(2), 619-644.
30. Schroeder, M. (2008). Number theory in science and communication: with applications in cryptography, physics, digital information, computing, and self-similarity (Vol. 7). Springer Science & Business Media.
31. Anderson, M., Banker, R., Huang, R., & Janakiraman, S. (2007). Cost behavior and fundamental analysis of SG&A costs. Journal of Accounting, Auditing & Finance, 22(1), 1-28.
32. Edwards, R. D., Magee, J., & Bassetti, W. H. C. (2007). Technical analysis of stock trends. CRC Press.
33. Bernstein, P. L. (1993). Capital ideas: The improbable origins of modern Wall Street. Simon and Schuster.
34. Ang, A., & Chen, J. (2007). CAPM over the long run: 1926–2001. Journal of Empirical Finance, 14(1), 1-40.
35. Lewellen, J., & Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. Journal of financial economics, 82(2), 289-314.
36. Gaunt, C. (2004). Size and book to market effects and the Fama French three factor asset pricing model: evidence from the Australian stockmarket.Accounting & Finance, 44(1), 27-44.

37. Mills, T. C., & Markellos, R. N. (2008). The econometric modelling of financial time series. Cambridge University Press.
38. Bauwens, L., & Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. Journal of Business & Economic Statistics, 23(3).
39. Mills, T. C., & Markellos, R. N. (2008). The econometric modelling of financial time series. Cambridge University Press.
40. Zivot, E., & Wang, J. (2007). Modeling Financial Time Series with S-PLUS®(Vol. 191). Springer Science & Business Media.
41. Chang, Ting-Cheng; Jane, Chuen-Jiuan & Lee, Yuan-Paio. (2006). A forecasting model of dynamic grey rough set and its application on stock selection. 2nd IEEE International Conference on Cybernetics & Intelligent Systems, 1-6.
42. Tay, Francis E.H. & Shen, Lixiang. (2002). Economic and financial prediction using rough sets model. European Journal of Operational Research, (141)3, 641-659.
43. Shen, Lixiang; Loh, Han Tong. (2004). Applying rough sets to market timing decisions. Decision Support Systems, (37)4, 583-597.
44. Tan, A. & Quek, C. (2005). Maximizing winning trades using a rough set based other-product (RSPOP) fuzzy neural network. IEEE Congress on Evolutionary, (3), 2076-2083.
45. Ang, Kai Keng & Quek, Chai. (2006). Stock trading using RSPOP: a novel rough set-based neuro-fuzzy approach. IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council, (17)5, 1301-1315.
46. **Zaini, Bahtiar Jamili;** Shamsuddin, Siti Mariyam & Jaaman, Saiful Hafizah. (2008). Predicting the Financial Performance of Publicly-Traded Malaysian Firms Using Rough Sets Based Feature Selection Techniques. The Icfai Journal of Applied Finance, (14)6.
47. Huang, Kuang Yu; Jane, Chuen-Jiuan & Chang, Ting-Cheng. (2008). A RS model for stock market forecasting and portfolio selection allied with weight clustering and grey system. IEEE Congress on Evolutionary, 1240-1246.
48. Huang, Kuang Yu et al. (2008). A new grey relation Analysis applied to the assert allocation of stock portfolio. International Journal of Computational Cognition, (6)3.
49. Huang, Kuang Yu & Jane, Chuen-Jiuan. (2008). A novel model for stock portfolio based on X, RS and a new grey relational grade theories. IEEE Confrence on Cybernetics, 417-422.
50. Huang, Kung-Yu & Jane, Cheuen- Jiuan. (2008). An automatic stock market forecasting and portfolio selection mechanism based on VPRS, ARX and grey system. IEEE Asian-Pacific Services Computing Conference, 4030-4035.
51. Huang, Kuang Yu. (2009). Application of VPRS model with enhanced threshold parameter selection mechanism to automatic stock market forecasting and portfolio selection. Expert Systems with Applications, (36)9, 11652-11661.
52. Huang, Kuang Yu & Jane, Chuen-jiuan. (2009). A hybrid model for stock market forecasting and portfolio selection based on ARX, Grey system and Rough set theories. Expert Systems with Applications, (36)3, 5387-5392.

53. Huang, Kuang Yu & Wan, Shiuan. Application of enhanced cluster validity index fuction to automatic stock portfolio selection system. Information technology & management, (12)3, 213-228.
54. Yao, JingTao & Herbert, Joseph P. (2009). Financial time-series analysis with rough sets. Applied Soft Computing, (9)3, 1000-1007.
55. Jaaman, Saiful Hafizah et al. (2009). A predictive model construction applying rough set methodology for Malaysian stock market return, International Research Journal of Finance and Economics, (30), 211-218.
56. Shamsuddin, Siti Mariyam; Jaaman, Saiful Hafizah & Darus, Maslina. (2009). Neuro-Rough Trading Rules for Mining Kuala Lumpur Composite Index. European Journal of Scientific Research, (28)2, 278-286.
57. Al-Qaheri, Hameed; Hassanien, Aboul Ella & Abraham, Ajith. (2009). A generic scheme for generating prediction rules using rough set. Rough set Theory: A True Landmark in Data Analysis, Studies in Computational Intelligence. (174), 163-186.
58. Chaudhuri, Arindam; De, Kajal & Chatterjee. Dipak. (2009). Discovering stock price prediction rules of Bambay Stock Exchange using rough fuzzy multi layer perception networks. Book Chapter: Forecasting Financial Markets in India, Indian Institute of Technology Kharagpur, Allied Publishers.
59. Cheng, Jao-Hong; Chen, Huei-Ping; Lin, Yi-Min. (2010). A hybrid forecast marketing timing model based on probabilistic neural network, rough set and C4.5. Expert System with Application, (37)3, 1814-1820.
60. Cheng, Ching-Hsue; Chen, Tai-Liang; Wei, Liang-Ying. (2010). A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. Information Sciences, (180)9, 1610-1629.
61. Banik, Shipra; Khodadad Khan, A F M & Anwer, Mohammad. (2014). Hybrid machine learning technique for forecasting Dhaka stock market timing decisions. Computational intelligence and neuroscience.
62. Saeid Hassan, Naglaa Ragaa. (2014). Forecasting the Egyptian's Market Trend Using Rough Set. International Journal of Advanced Research in Computer Science and Software Engineering, (4)1, 377-382.
63. Kirkdos S. & Manolopoulos Y., (2004), 'Data Mining in Finance and Accounting: A Review of Current Research Trends', Proceedings of the 1st International Conference on Enterprise Systems and Accounting (ICESAcc), Thessaloniki,Greece, pp. 63-78.
64. Gupta, G. K. (2014). Introduction to data mining with case studies. PHI Learning Pvt. Ltd..
65. Zaki, M. J., & Meira Jr, W. (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press.
66. Kudyba, S. (2014). Big Data, Mining, and Analytics: Components of Strategic Decision Making. CRC Press.
67. David Hand, Heikki Mannila , Padhraic Smyth. Principles of Data Mining. The MIT Press . 2001.
68. Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From Data Mining to Knowledge Discovery in Databases.1996
69. Camara, M. S., Naguingar, D., & Bah, A. (2015). Prior Data Quality Management in Data Mining Process. In New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering (pp. 299-307). Springer International Publishing.

70. Hassani, H., Saporta, G., & Silva, E. S. (2014). Data Mining and Official Statistics: The Past, the Present and the Future. Big Data, 2(1), 34-43.
71. Pálka, L., Schauer, F., & Zelinka, I. (2014). Artificial Intelligence Elements in Data Mining from Remote Experiments. In Nostradamus 2014: Prediction, Modeling and Analysis of Complex Systems (pp. 421-428). Springer International Publishing.
72. Ivezić, Ž., Connolly, A. J., VanderPlas, J. T., & Gray, A. (2014). Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. Princeton University Press.
73. Tayyebi, A., & Pijanowski, B. C. (2014). Modeling multiple land use changes using ANN, CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data mining tools. International Journal of Applied Earth Observation and Geoinformation, 28, 102-116.
74. Zhong, M., Schuetter, J., Mishra, S., & Lafollette, R. F. (2015, February). Do Data Mining Methods Matter?: A Wolfcamp Shale Case Study. In SPE Hydraulic Fracturing Technology Conference. Society of Petroleum Engineers.
75. Ko, H., Sung, J., Cho, S. M., & Yoon, T. (2014). Comparison of the Performances of the Decision Tree Algorithm C5 using Rough Set and the Neural Network.
76. Słowiński, R., Greco, S., & Matarazzo, B. (2014). Rough-set-based decision support. In Search Methodologies (pp. 557-609). Springer US.
77. Albanese, A., Pal, S. K., & Petrosino, A. (2014). Rough sets, kernel set, and spatiotemporal outlier detection. Knowledge and Data Engineering, IEEE Transactions on, 26(1), 194-207.
78. Jia, X., Tang, Z., Liao, W., & Shang, L. (2014). On an optimization representation of decision-theoretic rough set model. International Journal of Approximate Reasoning, 55(1), 156-166.
79. Yang, X. B., Song, X. N., Chen, Z. H., & Yang, J. Y. (2012). Incomplete Information System and Rough Set Theory (pp. 195-222). Berlin, Germany: Springer.
80. Lingras, P., Wolski, M., Cornelis, C., Mitra, S., & Wasilewski, P. (2013). Rough Sets and Knowledge Technology: 8th International Conference, RSKT 2013, Halifax, NS, Canada, October 11-14, 2013, Proceedings. Springer Publishing Company, Incorporated.
81. Polkowski, L. (2013). Rough sets: Mathematical foundations (Vol. 15). Springer Science & Business Media.
82. Podsiadło, M., & Rybiński, H. (2014). Rough Sets in Economy and Finance. InTransactions on Rough Sets XVII (pp. 109-173). Springer Berlin Heidelberg.
83. Yao, J., Yang, Y., Slowiński, R., Greco, S., Li, H., Mitra, S., & Polkowski, L. (Eds.). (2012). Rough Sets and Current Trends in Computing: 8th International Conference, RSCTC 2012, Chengdu, China, August 17-20, 2012. Proceedings(Vol. 7413). Springer.
84. Chakraborty, M. K., & Banerjee, M. (2013). Rough sets: Some foundational issues. Fundamenta Informaticae, 127(1), 1-15.
85. García, S., Luengo, J., & Herrera, F. (2015). Discretization. In Data Preprocessing in Data Mining (pp. 245-283). Springer International Publishing.
86. Kantardzic, M. (2011). Data mining: concepts, models, methods, and algorithms. John Wiley & Sons.
87. Castillo, J. E., & Miranda, G. F. (2013). Mimetic discretization methods. CRC Press.
88. Xiao, F., & Fan, C. (2014). Data mining in building automation system for improving building operational performance. Energy and buildings, 75, 109-118.

89. A. Øhrn, (2001). “ROSETTA Technical Reference Manual”. Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway. http://www.idi. ntnu.no/~aleks/rosetta/>
90. Hertzberg, M. P. (2013). Entanglement entropy in scalar field theory. Journal of Physics A: Mathematical and Theoretical, 46(1), 015402.
91. Adimurthi, Ghoshal, S. S., & Veerappa Gowda, G. D. (2012). STRUCTURE OF ENTROPY SOLUTIONS TO SCALAR CONSERVATION LAWS WITH STRICTLY CONVEX FLUX. Journal of Hyperbolic Differential Equations, 9(04), 571-611.
92. Grefenstette, J. J. (Ed.). (2013). Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms. Psychology Press.
93. Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. Statistics and Computing, 21(2), 137-146.
94. Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012). The ‘K’in K-fold Cross Validation. In European symposium on artificial neural networks.
95. Hornung, R., Bernau, C., Truntzer, C., Stadler, T., & Boulesteix, A. L. (2014). Full versus incomplete cross-validation: measuring the impact of imperfect separation between training and test sets in prediction error estimation.

Abstract:

Playing an important role in asset allocating, Stock Market gains lots of attention and tries to move caught funds in to producing industries. There have been many researches about forecasting stock return with different economic indicators. The models which designed for forecasting stock return, such as single factor and multi factor ones, where all linear. During the time researchers have found that the stock return behavior follows chaos theory and it is not linear, so soft computing and data mining techniques have been being started to be popular. The most important point for investors is finding the stock return behavior rules for deciding in more certainty environment and determining the appropriate time for buying and selling stocks. So using data mining techniques could be very helpful in this region. In this research The Rough Set Theory is being used. This method uses historical data for finding rules, and predicts the future behavior with those rules. Whats more the liquidity of the stock is being measured by means of fuzzy rough set theory and Adjusted Iliiquidity Amihud Measure.

Shahrood University of Technology

Faculty: Industry & Management Collogue

Determining systematic patterns in stock return by means of liquidity's return wave

Azadeh Hajjar

Supervisor:

Dr. Reza Sheikh

Advisor:

Dr. Seyed Mojtaba Mirlohi

February 2015