فصلنامه مهندسی مالی و مدیریت اوراق بهادار
دوره چهارده، شماره پنجاه و شش، پائیز ۱۴۰۲
نوع مقاله: علمی پژوهشی
صفحات: ۱۵۹-۱۴۲

شناسایی اندیکاتورهای موثر بر پیش‌بینی روند شاخص‌کل بورس اوراق بهادار تهران با استفاده از الگوریتم‌های انتخاب ویژگی و دسته‌بندی

محمدجواد شیخ‌زاده[۱]
سجاد رحمانی[۲]

تاریخ دریافت مقاله : ۱۴۰۱/۰۳/۱۹ تاریخ پذیرش مقاله : ۱۴۰۱/۰۶/۲۱

چکیده

پیش‌بینی روند تغییرات شاخص و قیمت سهام مسئله بسیار پیچیده‌ای است که تحت تاثیر عوامل مختلف محیطی، صنعتی، خرد یا کلان درون شرکتی قرار دارد. اندازه‌گیری عوامل ذکر شده بسیار مشکل است و یا در بعضی موارد قابلیت اندازه گیری ندارند، ولی برخی از عوامل خرد مانند سابقه قیمت و حجم معاملات به سادگی قابل اندازه‌گیری و در دسترس است. در این پژوهش با استفاده از الگوریتم‌های انتخاب ویژگی و دسته‌بندی، روشی جهت تعیین موثرترین اندیکاتورها برای شاخص‌کل و شاخص‌کل هم‌وزن ارائه خواهد شد. سپس مدل پیشنهادی را در بازه زمانی روزانه و هفتگی از دی ماه ۱۳۹۸ الی اردیبهشت ماه ۱۴۰۱ مورد بررسی قرار خواهیم داد. نتایج نشان می‌دهد پیش‌بینی روند تغییرات با استفاده از تعداد محدودی از اندیکاتورها با دقت بالایی امکان‌پذیر است و همچنین در بازه زمانی هفتگی نسبت به بازه زمانی روزانه دقت پیش‌بینی به طور قابل محسوسی بیشتر می‌باشد.

کلمات کلیدی

شاخص‌کل، اندیکاتور، انتخاب ویژگی، دسته‌بندی

۱-گروه علوم کامپیوتر، دانشکده ریاضی و علوم کامپیوتر، دانشگاه دامغان، دامغان، ایران. Sheikhzadeh@gonbad.ac.ir
۲-گروه علوم کامپیوتر، دانشکده ریاضی و علوم کامپیوتر، دانشگاه دامغان، دامغان، ایران. (نویسنده مسوول) Rahmani@du.ac.ir

مقدمه

بازار سهام به دلیل پتانسیل کسب سود در یک دوره زمانی کوتاه نسبت به بازارهای دیگر، یکی از جذابترین بازارهای مالی جهت سرمایه‌گذاری می‌باشد. از این رو پیش‌بینی تغییرات قیمت در بازار سهام برای سرمایه‌گذاران بسیاری با اهمیت و حیاتی است. طبیعت غیرخطی، پویا و ناپایدار داده‌های سهام باعث شده ایجاد یک سیستم مطمئن با قابلیت پیش‌بینی دقیق قیمت سهام بسیار دشوار باشد. به دلیل تغییرات نامنظم، داده‌های بازار سهام بسیار پیچیده تر از داده‌های آماری است. عوامل بسیاری همچون شرایط سیاسی، اقتصادی، فصلی و روندها بر روی قیمت سهام تاثیرگذار هست [۶]. برای پیش‌بینی بازار سهام روش‌های زیادی وجود دارد که به مدل‌های آماری و مدل‌های یادگیری ماشین تقسیم می‌شوند. مدل‌های آماری خطی بوده و پیاده‌سازی آنها ساده است. اما این مدل‌ها قادر به شناسایی اطلاعات مخفی موجود در ساختار غیرخطی داده‌های بازار سهام نیستند [۹]. اخیراً در بسیاری از مسائل آماری از روش‌های هوش مصنوعی و یادگیری ماشین استفاده می‌شود که یکی از مهمترین آنها، پیش‌بینی بازارهای مالی از جمله بازار سهام است. از ویژگی‌های این روش‌ها می‌توان به غیرخطی بودن، خودآموزی، خودانطباقی و خودسازماندهی بودن آنها اشاره کرد، که باعث می‌شود تا بتوانند اطلاعات مخفی موجود در داده‌های ورودی را حتی در مواردی که کشف روابط تابعی بسیار مشکل است، شناسایی کنند. از جمله این الگوریتم‌ها می‌توان به شبکه عصبی پس‌انتشار[۱]، شبکه عصبی مصنوعی پیوند عملکردی[۲]، شبکه عصبی موجک[۳]، شبکه عصبی بازگشتی[۴]، نزدیکترین همسایه محدود به شعاعی[۵]، بهینه سازی گرگ خاکستری- شبکه عصبی المان[۶]، تجزیه و تحلیل مؤلفه‌های اصلی هسته[۷]، ماشین بردار پشتیبان[۸] و نیو بیز[۹] اشاره کرد [۲۱، ۲۴، ۱۶، ۱۵، ۶، ۲۳ و ۱۴].

تحلیل تکنیکال پیش‌بینی روند قیمت آتی بر اساس بررسی روند قیمت گذشته است. با استفاده از این روش نمی‌توان روند آینده قیمت را بطور مطلق پیش‌بینی کرد، بلکه می‌توان به سرمایه‌گذاران کمک کرد تا پیش‌بینی کنند که در طول زمان چه اتفاقی ممکن است برای قیمت‌ها رخ دهد. همچنین تحلیل تکنیکال با عوامل موثر بر عرضه و تقاضای سهام مرتبط است که به درک ارزش ذاتی سهام و دانستن اینکه آیا سهام کمتر یا بیشتر ارزش‌گذاری شده‌اند کمک می‌کند. اندیکاتورهای بازار سهام به سرمایه‌گذار کمک می‌کنند تا نقاط عطف بازار را شناسایی کند. تحلیل تکنیکال را می‌توان به عنوان هنر و علم پیش‌بینی قیمت‌های آینده بر اساس بررسی تغییرات قیمت‌ها در گذشته تعریف کرد. همچنین می‌توان گفت که تحلیل تکنیکال عبارتست از قرار دادن اطلاعات سهام مانند قیمت‌ها و حجم‌ها در نمودار و اعمال الگوها و شاخص‌های مختلف برای ارزیابی تغییرات قیمت در آینده. بازه زمانی که در آن تحلیل

تکنیکال اعمال می‌شود ممکن است از داده‌های قیمت دقیقه‌ای، ساعتی، روزانه، هفتگی یا ماهانه تا چندین‌ساله متغیر باشد.

اساساً دو روش برای تجزیه و تحلیل فرصت‌های سرمایه‌گذاری در بازار وجود دارد که عبارتند از تحلیل بنیادی و تحلیل تکنیکال. تحلیل بنیادی بر اطلاعات بنیادی مانند جنبه‌های مالی و غیر مالی شرکتها تکیه دارد در حالی که تحلیل تکنیکال اصول بنیادی را نادیده می‌گیرد و بر تغییرات واقعی قیمت تمرکز دارد. تحلیلگر تکنیکال فرض می‌کند که ۹۰ درصد معاملات روانی و ۱۰ درصد منطقی هستند. تحلیلگران تکنیکال سعی نمی‌کنند ارزش ذاتی اوراق بهادار را اندازه‌گیری کنند، بلکه از نمودارها و ابزارهای دیگر برای شناسایی الگوهایی استفاده می‌کنند که می‌توانند فعالیت‌های آینده را پیشنهاد بدهند. مفهوم تحلیل زمانی مطرح می‌شود که باید در مورد انتخاب سهام یک شرکت خاص برای سرمایه‌گذاری تصمیم‌گیری شود. تجزیه و تحلیل تکنیکال یک تکنیک تجزیه و تحلیل اوراق بهادار است که ادعا می‌کند توانایی پیش‌بینی جهت آینده قیمت‌ها از طریق مطالعه داده‌های گذشته بازار را دارد (عمدتاً قیمت و حجم معاملات).

مفروضات اساسی تحلیل تکنیکال شامل موارد زیر است:

۱- نوسانات بازار خلاصه‌ای از هر چیز دیگری است.

تحلیلگران تکنیکال بر این باورند که تغییرات در قیمت و نحوه معامله یک اوراق بهادار در بازار، تمام اطلاعات موجود در مورد آن را شامل می‌شود، بنابراین نشان دهنده ارزش منصفانه آن اوراق بهادار است. یعنی هرعامل بنیادی که می‌تواند برروی یک شرکت تاثیر بگذارد، خودش را در قیمت سهام نشان می‌دهد. به‌همین دلیل تحلیلگران تکنیکال معمولا به عواملی که برای تحلیلگران‌بنیادی مهم است، توجهی ندارند.

۲- حرکات قیمت را اغلب می‌توان با نمودار ترسیم و پیش‌بینی کرد.

تحلیلگران تکنیکال اذعان دارند که دوره‌های زمانی وجود دارد که قیمت‌ها به‌طور تصادفی حرکت می‌کنند، اما زمان‌هایی نیز وجود دارد که قیمت‌ها در یک روند قابل شناسایی حرکت می‌کنند. هنگامی که یک روند مشخص شد، می‌توان با معاملات هدفمند از آن درآمد کسب کرد. با تنظیم مدت زمانی که بازار در حال تجزیه و تحلیل است، می‌توان روندهای کوتاه مدت و بلندمدت را شناسایی کرد.

۳- تاریخ دوباره تکرار می‌شود.

مردم انگیزه‌های خود را یک شبه تغییر نمی‌دهند. بنابراین، زمانی که شرایطی تکرار می‌شود، معامله‌گران به همان شیوه‌ای که در گذشته واکنش نشان داده‌اند عمل می‌کنند. از آنجایی که مردم هر

بار واکنش مشابهی نشان می‌دهند، تحلیلگران تکنیکال می‌توانند از دانش نحوه واکنش معامله‌گران دیگر در گذشته استفاده کنند، تا هر بار که آن شرایط تکرار می‌شود، سود ببرند [۴، ۸، ۱۹ و ۲۰].

اصولا تحلیل تکنیکال برای بازارهایی مناسب است که قیمت در آن بر اساس عرضه و تقاضا تعیین می‌شود. بنابراین قیمت خودرو، بنزین، دارو و ... که در ایران به صورت دستوری قیمت‌گذاری می‌شود را نمی‌توان بر اساس تحلیل تکنیکال پیش‌بینی کرد زیرا هر گونه دستکاری در قیمت یک دارایی، بر روند طبیعی معاملات آن تاثیرگذار است. در بازار بورس ایران علیرغم اینکه در برخی از سهام دستکاری قیمت دیده می‌شود، اما بسیاری از سهام وجود دارند که از این امر مصون مانده‌اند و تحلیل تکنیکال در مورد آنها قابل استفاده است.

از نظر تاریخی، تحقیقات علمی کمتری بر روی تحلیل تکنیکال نسبت به تحلیل بنیادی صورت گرفته است [۲۰]. با وجود محبوبیت تحلیل تکنیکال در میان معامله‌گران، تا قبل از ۱۹۷۰ پژوهشگران تقریباً به اتفاق آرا معتقد بودند که این نوع تحلیل در عمل هیچ ارزش اقتصادی ندارد تا اینکه الکساندر [۱] به این نتیجه رسید که در بازارهای سوداگرانه ، حرکت قیمت، پس از شروع تمایل به تداوم دارد. با این حال او همچنین متوجه شد که کارمزدها می‌توانند سودآوری را کاهش دهند. سپس اشتباه محاسباتی تصحیح شده و هزینه‌های مبادله‌ای در نظر گرفته شد [۲]. فاما سعی کرد با آزمون‌های مختلف نشان دهد که تغییرات قیمت مستقل است و نمی‌توان از قیمت‌های تاریخی سهام برای پیش‌بینی‌های آینده به شکل ارزشمندی استفاده کرد [۱۲]. تقریبا تا سال ۲۰۰۰ در میان دانشگاهیان ارزش اقتصادی تحلیل تکنیکال همچنان بحث‌برانگیز باقی ماند [۲۸].

با این حال در سال‌های اخیر تحقیقات متعددی وجود دارد که اندیکاتورهای تکنیکال را به‌عنوان مکانیسم اصلی جهت پیش‌بینی قیمت سهام بکار گرفته‌اند. مطالعات متعددی با استفاده از الگوریتم‌های مختلف برای پیش‌بینی روند قیمت در بازارهای مالی انجام شده است، اما به دلیل ماهیت پیچیده و غیرخطی بودن رفتار بازار مالی، هنوز مشخص نیست که کدام الگوریتم می‌تواند عملکرد بهتری داشته باشد. لین و همکاران روش PCA را برای کاهش و فیلتر کردن نویز در داده‌ها پیشنهاد کردند [۱۷]. طالبی و همکاران روش طبقه بندی جدیدی را برای شناسایی روندهای صعودی، نزولی و جانبی نرخ ارز در بازار فارکس پیشنهاد دادند [۲۵]. زهونگ و انکه روش‌های داده‌کاوی را برای پیش‌بینی قیمت سهام به صورت روزانه مورد مطالعه قرار دادند. آنها ویژگی‌های مالی و اقتصادی مختلفی را معرفی کردند و سپس با استفاده از روش‌هایی مانند KPCA و FRPCA[۱۰] ویژگی‌ها را کاهش دادند [۲۹]. پاتل و همکاران روند قیمت سهام را با استفاده از ده شاخص فنی به عنوان ویژگی‌های ورودی الگوریتم‌های

شبکه عصبی مصنوعی مانند ماشین بردار پشتیبان، جنگل تصادفی و بیز ساده پیش‌بینی کردند [۲۲]. دش و دش از شش شاخص فنی اساسی به عنوان ویژگی برای یک سیستم پشتیبان تصمیم‌گیری مبتنی بر شبکه عصبی مصنوعی استفاده کردند [۱۱]. همچنین چن و همکاران کارایی روش یادگیری عمیق را با روش‌های یادگیری ماشینی مانند انتشار برگشتی[۱۱]، ماشین یادگیری شدید[۱۲] و تابع پایه شعاعی[۱۳] برای پیش‌بینی رفتار بازار سهام چین مقایسه کردند [۷].

اندیکاتورهای متعددی برای پیش‌بینی روند قیمت سهام استفاده می‌شود که بعضی از آنها علی‌رغم بار محاسباتی زیاد، دقت پایینی دارند. بنابراین پیدا کردن ترکیبی از اندیکاتورها که بتواند روند افزایش و یا کاهش قیمت سهام را پیش‌بینی کند همچنان حائز اهمیت است. استفاده از الگوریتم‌های انتخاب ویژگی می‌تواند به کاهش شاخص‌ها و ویژگی‌های اضافی، کاهش پیچیدگی و بهبود کارایی روش‌های پیش‌بینی بازارهای مالی کمک کند. اهداف انتخاب ویژگی را می‌توانیم به شکل زیر بیان کنیم:

۱- کاهش پیچیدگی محاسباتی داده‌های بزرگ با ارزیابی و تحلیل سخت، پرهزینه و زمان‌بر

۲- افزایش کارایی و درک بهتر پردازش‌های زیربنایی

۳- افزایش قدرت دسته‌بندی با افزودن ویژگی‌های حاوی اطلاعات مفید

از آنجایی که در تحلیل تکنیکال با یک پروژه و یا مسئله پیچیده و حجم زیادی از داده‌ها و متغیرها روبه رو هستیم و معادله و فرمول مشخصی با کارایی بالا برای حل این مسائل وجود ندارد، استفاده از روش‌های یادگیری ماشین ازجمله انتخاب ویژگی و دسته‌بندی برای افزایش کارایی تحلیل تکنیکال ضروری به نظر می‌رسد. همچنین برخی از اندیکاتورها باعث می‌شوند تحلیلگران بازار سهام در مورد روند قیمت سهام قضاوت نادرستی داشته باشند که با استفاده از روش‌های یادگیری ماشین و انتخاب مهم‌ترین اندیکاتورها برای تحلیل هر سهم کار تحلیلگر بسیار ساده خواهد شد، هدف این مقاله ارایه روشی برای تحلیلگران بازار سهام است تا از شاخص‌های مهم در هر سهم برای بهبود دقت پیش‌بینی استفاده کنند و همچنین با کاهش تعداد ویژگی‌ها، محاسبات را ساده‌تر و سریعتر انجام دهند. در مطالعات گذشته کمتر به انتخاب ویژگی‌ها و اندیکاتورهای موثر با استفاده از الگوریتم‌های یادگیری ماشین پرداخته شده، همچنین تعداد ویژگی‌ها و اندیکاتورهای مورد مطالعه در این مطالعات نسبتا اندک است و به تفاوت بین بازه زمانی روزانه و هفتگی در الگوریتم‌های یادگیری ماشین کمتر پرداخته شده است. در این مقاله از تعداد ۳۰ ویژگی و اندیکاتور استفاده شده، تمرکز بیشتر بر انتخاب ویژگی‌های مهم‌تر از بین این ۳۰ ویژگی جهت افزایش کارایی بوده و همچنین به بررسی تفاوت دقت دسته‌بندی داده‌های بازه زمانی روزانه با داده‌های بازه زمانی هفتگی پرداخته شده است.

در حقیقت موضوع کاهش تعداد ویژگی در بسیاری از کاربردها، مانند دسته‌بندی، حیاتی است، در عمل ویژگی‌های زیادی وجود دارند که بسیاری از آنها بی فایده یا فاقد اطلاعات هستند. حذف نکردن این ویژگی‌ها مشکل اطلاعاتی ایجاد نمی‌کند، اما محاسبات را پیچیده کرده و باعث می‌شوند اطلاعات زیادی در کنار داده‌های مفید ذخیره شوند. بهمین منظور در این پژوهش داده‌ها با استفاده از الگوریتم‌های مختلف انتخاب ویژگی مورد تجزیه و تحلیل قرار گرفتند که هر کدام زیر مجموعه‌ای از ویژگی‌های اولیه را انتخاب کردند. سپس از یک الگوریتم دسته‌بندی برای ارزیابی کیفیت ویژگی‌های انتخاب شده استفاده شد و در نهایت به این نتیجه رسیدیم که هرچه کیفیت ویژگی‌ها برای دسته‌بندی بالاتر باشد، دسته‌بندی با استفاده از آن ویژگی‌ها از دقت بیشتری برخوردار است.

در ادامه و در بخش ۲ داده‌ها و الگوریتم‌های مورد استفاده معرفی شده و یک نمای کلی از روش‌های ارزیابی کارایی نشان داده شده است. نتایج تجربی در بخش ۳ ارائه شده و بخش ۴ حاوی نتایج و توصیه‌هایی برای کارهای آینده است.

روش‌شناسی تحقیق

در این پژوهش تعداد ۳۰ ویژگی و اندیکاتور برای شاخص‌کل و شاخص‌کل هم‌وزن بورس اوراق بهادار تهران محاسبه و مورد استفاده قرار گرفت شامل: قیمت اولین معامله روز، قیمت آخرین معامله روز، بیشترین قیمت روز، کمترین قیمت روز، حجم معاملات روز، ZigZag، RSI، Macd، Signal، MFI، STOCHD، STOCHK، Momentum، Kijun-sen، Tenkan-sen، Chikou-Span، Senkou-Span A، Senkou-Span B، DEMA، DTOscSK، DTOscSD، NNRSI، OBV، MA(OBV)، Williams %R، CCI، Teeth، Jaw، Lips و نرخ دلار.

فرآیند انتخاب ویژگی بصورت بخش‌های زیر تقسیم‌بندی می‌شود [۱۰]:

۱- روش جستجو[۱۴]: تابع تولید کننده زیرمجموعه‌های کاندید شده را برای روش مورد نظر پیدا می‌کند.

۲- مدل ارزیابی[۱۵]: زیر مجموعه نامزد شده را بر اساس روش جستجو ارزیابی کرده و یک عدد به عنوان میزان مزیت روش باز می‌گرداند. روش‌های مختلف تلاش می‌کنند تا زیرمجموعه‌ای را پیدا کنند که این مقدار را بهینه کند.

۳- شرط خاتمه : برای تصمیم‌گیری در مورد زمان توقف الگوریتم استفاده می‌شود.

جهت کاهش تعداد ویژگی‌ها از الگوریتم‌های انتخاب ویژگی موجود در کتابخانه Weka استفاده شد. روش‌های جستجوی Bestfirst و Attribute Ranking، همچنین مدل‌های ارزیابی مختلف از جمله: ارزیاب زیر مجموعه انتخاب ویژگی مبتنی بر همبستگی[۱۶]، ارزیاب ویژگی همبستگی[۱۷]، ارزیاب ویژگی کسب اطلاعات[۱۸]، ارزیابی ویژگی نسبت بهره[۱۹] و ارزیاب ویژگی متقارن[۲۰] [۵].

برای ارزیابی کیفیت ویژگی‌های انتخاب شده، از سه روش دسته‌بندی استفاده شد که از جمله پرکاربردترین الگوریتم‌های دسته‌بندی هستند: جنگل تصادفی[۲۱] [۳]، الگوریتم C4.5 [۲۶] و ماشین بردار پشتیبان[۲۲] [۲۶]. برای اعتبارسنجی این الگوریتم‌ها نیز از روش 10-Fold Cross-Validation استفاده شد [۵]. بدین منظور نرم‌افزاری مبتنی بر جاوا طراحی و پیاده‌سازی شد که تاریخچه معاملات روزانه شاخص‌کل و شاخص‌کل هم‌وزن را به طور برخط از سایت شرکت مدیریت فناوری بورس تهران دریافت می‌کند، اندیکاتورها را محاسبه کرده، عملیات پردازشی مانند انتخاب ویژگی و دسته‌بندی روی داده‌های نرمال شده اعمال شده و در نهایت ویژگی‌های انتخابی و خطاهای دسته‌بندی را برای هر نماد به صورت یک فایل اکسل ذخیره می‌کند. در ادامه نرم‌افزار فوق برای پردازش داده‌های تاریخچه معاملات روزانه دو سهم فاسمین و بپاس نیز مورد استفاده قرار گرفت.

داده‌های مورد استفاده در این تحقیق شامل تاریخچه معاملات روزانه شاخص‌کل هم وزن و شاخص‌کل و همچنین تاریخچه معاملات روزانه دو نماد فاسمین و بپاس از دی ماه ۱۳۹۸ تا اردیبهشت ۱۴۰۱ و ۵۴۵ روز کاری است، که از سایت شرکت مدیریت فناوری بورس اوراق بهادار تهران به آدرس http://tsetmc.com دریافت شده است.

جهت ارزیابی کارایی الگوریتم‌ها از معیارهای متعددی استفاده شد که در ادامه معرفی می‌شوند:

میانگین خطای مطلق، میانگین قدرمطلق تفاضل بین مقدار پیش‌بینی شده و مقدار واقعی را محاسبه می‌کند [۶].

$$\mathrm{MAE} = \frac{1}{N}\sum_{k=1}^{n} |A_k - P_k| \qquad (۱)$$

ریشه میانگین مربعات خطا، ریشه میانگین مربعات فاصله بین مقدار پیش‌بینی شده ومقدار واقعی را محاسبه می‌کند [۶].

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{n}(|A_k - P_k|)^2} \qquad (۲)$$

دقت دسته‌بندی برابر است با نسبت تعداد نمونه‌های درست پیش‌بینی شده به تعداد کل پیش‌بینی‌های انجام شده و ضرب نتیجه در ۱۰۰

$$\mathrm{ACC} = \frac{T}{N} * 100 \qquad (۳)$$

جایی که T تعداد نمونه‌هایی است که درست دسته‌بندی شده‌اند و N تعداد کل نمونه‌ها است.

Sensitivity نشان دهنده درصد کلاس‌های مثبتی است که به درستی دسته‌بندی شده‌اند [۲۷].

$$S_n = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad (۴)$$

Specificity نشان می‌دهد که چه مقدار از کلاس‌های منفی به درستی دسته‌بندی شده است [۲۷]:

$$S_p = \frac{\mathrm{T}N}{\mathrm{TN+FP}} \qquad (۵)$$

AUC-ROC: منحنی مشخصه اپراتور گیرنده (ROC) یک معیار ارزیابی برای مسائل دسته‌بندی باینری است. یک منحنی احتمال که TPR را در برابر FPR در مقادیر آستانه مختلف ترسیم کرده و اساساً "سیگنال" را از "نویز" جدا می‌کند. مساحت زیر منحنی (AUC) اندازه‌گیری توانایی دسته‌بندی برای تمایز بین کلاس‌ها است و به عنوان خلاصه منحنی ROC گفته می‌شود. هر چه AUC بیشتر باشد، عملکرد مدل در تشخیص کلاس‌های مثبت و منفی بهتر است [۱۳]:

$$AUC = \frac{1+S_n-FP_{rate}}{2} \qquad (۶)$$

جایی که FP_{rate} نسبت تعداد مثبت کاذب به تعداد کل روزهای غیر پرت است:

$$FP_{rate} = \frac{\mathrm{FP}}{\mathrm{TN+FP}} = 1 - S_p \qquad (۷)$$

ضریب همبستگی متیو[۲۳] تعادل بهتری را بین چهار معیار اصلی S_n، S_p، PPV و NPV فراهم می‌کند [۱۸]:

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TN+FN)*(TP+FP)*(TP+FN)*(TN+FP)}} \qquad (۸)$$

یافته‌های پژوهش

جداول ۱ تا ۳ نتایج الگوریتم‌های انتخاب ویژگی و دسته‌بندی مورد استفاده بر روی داده‌ها را نشان می‌دهد. این یافته‌ها بیانگر آن است که با تعداد اندکی از اندیکاتورها و ویژگی‌ها، می‌توان روند معاملات آتی را با دقت مناسب و در برخی موارد با دقت بیشتر از زمانی که از همه ۳۰ ویژگی استفاده می‌شود، پیش‌بینی کرد. به عنوان مثال، تنها با ۱۰ ویژگی (Senkou-Span A، RSI، نرخ دلار، MAOBV، Kijun-Sen، OBV، CCI، ZigZag، Williams %R و Chikou-Span)، دقت پیش‌بینی روند شاخص‌کل با استفاده از الگوریتم انتخاب ویژگی Gain Ratio-Attribute Ranking و با روش دسته‌بندی جنگل تصادفی در دوره زمانی هفتگی ۹۱,۴۵۵ درصد شد، در حالی که بدون الگوریتم انتخاب

ویژگی و با استفاده از تمامی ۳۰ ویژگی، دقت پیش‌بینی ۹۰,۳۶۴ درصد بود. شکل ۱ فراوانی ویژگی‌های انتخاب شده توسط الگوریتم‌های انتخاب ویژگی را نشان می‌دهد و بیانگر آن است که اندیکاتورهای RSI، Jaw، Senkou-Span A، CCI، Williams %R، Senkou-Span B، Teeth و OBV بیشتر نسبت به سایر اندیکاتورها و ویژگی‌ها توسط الگوریتم‌های انتخاب ویژگی انتخاب شده‌اند، بنابراین احتمالا می‌توانند در تشخیص روند شاخص‌کل و شاخص‌کل هم‌وزن نسبت به سایر ویژگی‌ها و اندیکاتورها موثرتر باشند. نتایج دسته‌بندی این هشت اندیکاتور با الگوریتم دسته‌بندی جنگل تصادفی در جدول ۴ نشان داده شده است. با توجه به این نتایج، تنها با محاسبه هشت اندیکاتور مذکور می‌توان روند تغییرات شاخص‌ها را با دقت مناسب پیش‌بینی کرد.

علاوه بر شاخص‌کل و شاخص‌کل هم‌وزن، دسته‌بندی ۳۰ ویژگی اولیه و هشت اندیکاتور منتخب با الگوریتم جنگل تصادفی برای نمادهای فاسمین و بپاس نیز انجام شد، که نتایج آن در جدول ۵ نشان داده شده است، این نتایج بیانگر آن است که هشت اندیکاتور منتخب برای این دو نماد هم با دقت مناسبی دسته‌بندی شده‌اند و حتی در مورد این دو نماد دقت دسته‌بندی با هشت اندیکاتور منتخب بیشتر از دسته‌بندی با ۳۰ اندیکاتور اولیه است.

کد جاوا و داده‌های مرتبط با این تحقیق از طریق آدرس زیر قابل دسترس است:
http://faculty.du.ac.ir/rahmani

جدول ۱: نتایج الگوریتم‌های انتخاب ویژگی و دسته‌بندی جنگل تصادفی

نماد- دوره زمانی	روش جستجو- مدل ارزیابی	ACC	MAE	RMSE	تعداد ویژگی‌های منتخب	AUC-ROC	MCC
شاخص‌کل هم وزن- روزانه	CFS- BEST FIRST	۶۸.۵۰۸	۰,۳۷۲	۰.۴۴۹	۱۳	۰,۷۵۹	۰,۳۶۷
شاخص‌کل هم وزن- هفتگی	CFS- BEST FIRST	۹۰,۹۷۶	۰,۱۶۱	۰,۲۵۵	۱۲	۰,۹۷۳	۰,۸۱۹
شاخص‌کل- روزانه	CFS- BEST FIRST	۵۹.۴۵۵	۰.۴۴۳	۰.۵۰۴	۵	۰.۶۲۰	۰,۱۷۱
شاخص‌کل- هفتگی	CFS- BEST FIRST	۸۸,۷۲۷	۰,۱۷۷	۰,۲۷۶	۵	۰,۹۶۴	۰,۷۶۷
شاخص‌کل هم وزن- روزانه	InfoGain- Attribute Ranking	۶۸.۵۰۸	۰,۳۷۸	۰.۴۵۸	۱۳	۰,۷۴۵	۰,۳۶۶
شاخص‌کل هم وزن- هفتگی	InfoGain- Attribute Ranking	۹۱,۳۴۴	۰,۱۵۴	۰,۲۴۹	۲۱	۰,۹۷۴	۰,۸۲۶
شاخص‌کل- روزانه	InfoGain- Attribute Ranking	۵۹.۴۵۵	۰.۴۴۰	۰.۵۰۴	۷	۰.۶۲۵	۰,۱۷۰
شاخص‌کل- هفتگی	InfoGain- Attribute Ranking	۸۹.۶۳۶	۰,۱۷۳	۰,۲۶۹	۱۱	۰,۹۶۸	۰,۷۸۶
شاخص‌کل هم وزن- روزانه	GainRatio- Attribute Ranking	۶۷.۵۸۸	۰,۳۷۵	۰.۴۵۵	۱۰	۰,۷۴۹	۰,۳۴۸
شاخص‌کل هم وزن- هفتگی	GainRatio- Attribute Ranking	۹۲,۲۶۵	۰,۱۶۱	۰,۲۴۹	۳۰	۰,۹۷۵	۰,۸۴۵
شاخص‌کل- روزانه	GainRatio- Attribute Ranking	۶۲,۳۶۴	۰.۴۳۰	۰.۴۸۹	۳۰	۰.۶۵۴	۰,۲۲۶
شاخص‌کل- هفتگی	GainRatio- Attribute Ranking	۹۱.۴۵۵	۰,۱۹۰	۰,۲۷۸	۱۰	۰,۹۶۳	۰,۸۲۵

شاخص‌کل هم وزن- روزانه	Correlation- Attribute Ranking	۶۵,۵۶۲	۰,۳۹۴	۰.۴۷۲	۸	۰,۷۱۸	۰,۳۰۶
شاخص‌کل هم وزن- هفتگی	Correlation- Attribute Ranking	۹۰,۲۳۹	۰,۱۶۹	۰,۲۶۶	۱۰	۰,۹۶۸	۰,۸۰۴
شاخص‌کل- روزانه	Correlation- Attribute Ranking	۶۰,۳۶۴	۰.۴۴۰	۰.۵۱۰	۹	۰.۶۱۸	۰,۱۸۸
شاخص‌کل- هفتگی	Correlation- Attribute Ranking	۸۸,۹۰۹	۰,۱۷۸	۰,۲۷۵	۱۰	۰,۹۶۴	۰,۷۷۱
شاخص‌کل هم وزن- روزانه	Symmetrical- Attribute Ranking	۶۶,۶۶۷	۰,۳۷۸	۰.۴۵۸	۱۲	۰,۷۴۵	۰,۳۲۸
شاخص‌کل هم وزن- هفتگی	Symmetrical- Attribute Ranking	۹۲,۲۶۵	۰,۱۵۴	۰,۲۵۱	۱۸	۰,۹۷۲	۰,۸۴۵
شاخص‌کل- روزانه	Symmetrical- Attribute Ranking	۶۱,۰۹۱	۰.۴۳۴	۰.۴۹۴	۱۰	۰.۶۴۰	۰,۲۰۲
شاخص‌کل- هفتگی	Symmetrical- Attribute Ranking	۸۹.۶۳۶	۰,۱۸۰	۰,۲۷۴	۹	۰,۹۶۵	۰,۷۸۶
شاخص‌کل هم وزن- روزانه	No Filter	۷۰,۱۶۶	۰,۳۶۴	۰.۴۴۵	۰	۰,۷۷۰	۰.۴۰۱
شاخص‌کل هم وزن- هفتگی	No Filter	۹۲,۲۶۵	۰,۱۶۱	۰,۲۴۹	۰	۰,۹۷۵	۰,۸۴۵
شاخص‌کل- روزانه	No Filter	۶۲,۳۶۴	۰.۴۳۰	۰.۴۸۹	۰	۰.۶۵۴	۰,۲۲۶
شاخص‌کل- هفتگی	No Filter	۹۰,۳۶۴	۰,۱۹۴	۰,۲۷۸	۰	۰,۹۶۵	۰,۸۰۱

(منبع: یافته‌های پژوهشگر)

جدول ۲: نتایج الگوریتم‌های انتخاب ویژگی و دسته‌بندی C4.5

نماد- دوره زمانی	روش جستجو- مدل ارزیابی	ACC	MAE	RMSE	تعداد ویژگی‌های منتخب	AUC-ROC	MCC
شاخص‌کل هم وزن- روزانه	CFS- BEST FIRST	۶۴,۶۴۱	۰.۴۰۳	۰.۴۸۹	۱۳	۰.۶۸۹	۰,۲۸۴
شاخص‌کل هم وزن- هفتگی	CFS- BEST FIRST	۸۶.۳۷۲	۰,۱۵۵	۰,۳۵۹	۱۲	۰,۸۶۲	۰,۷۲۶
شاخص‌کل- روزانه	CFS- BEST FIRST	۵۸,۰۰۰	۰.۴۶۶	۰.۵۰۳	۵	۰.۵۷۰	۰,۱۰۳
شاخص‌کل- هفتگی	CFS- BEST FIRST	۸۱,۲۷۳	۰,۲۳۴	۰,۳۸۸	۵	۰,۸۵۴	۰.۶۱۷
شاخص‌کل هم وزن- روزانه	InfoGain- Attribute Ranking	۶۳,۹۰۴	۰.۴۱۷	۰.۴۹۳	۱۳	۰.۶۷۳	۰,۲۶۹
شاخص‌کل هم وزن- هفتگی	InfoGain- Attribute Ranking	۸۸,۲۱۴	۰,۱۳۴	۰,۳۳۲	۲۱	۰,۸۸۷	۰,۷۶۳
شاخص‌کل- روزانه	InfoGain- Attribute Ranking	۵۶,۵۴۶	۰.۴۷۲	۰.۵۰۱	۷	۰.۵۷۹	۰,۱۶۵
شاخص‌کل- هفتگی	InfoGain- Attribute Ranking	۸۵.۸۱۸	۰,۱۶۲	۰,۳۵۴	۱۱	۰,۸۸۹	۰,۷۰۹
شاخص‌کل هم وزن- روزانه	GainRatio- Attribute Ranking	۶۴,۶۴۱	۰.۴۱۵	۰.۴۸۳	۱۰	۰.۶۷۹	۰,۲۸۵
شاخص‌کل هم وزن- هفتگی	GainRatio- Attribute Ranking	۸۷,۸۴۵	۰,۱۴۰	۰,۳۳۹	۳۰	۰,۸۷۸	۰,۷۵۶
شاخص‌کل- روزانه	GainRatio- Attribute Ranking	۵۸,۰۰۰	۰.۴۵۹	۰.۵۲۳	۳۰	۰.۵۸۲	۰,۱۱۶
شاخص‌کل- هفتگی	GainRatio- Attribute Ranking	۸۵.۰۹۱	۰,۱۷۹	۰,۳۶۱	۱۰	۰,۸۷۷	۰,۷۰۰
شاخص‌کل هم وزن- روزانه	Correlation- Attribute Ranking	۶۶.۱۱۴	۰.۴۱۶	۰.۴۷۸	۸	۰.۶۸۰	۰,۳۱۴
شاخص‌کل هم وزن- هفتگی	Correlation- Attribute Ranking	۸۷,۲۹۳	۰,۱۶۱	۰,۳۴۰	۱۰	۰,۸۸۲	۰,۷۴۵
شاخص‌کل- روزانه	Correlation- Attribute Ranking	۵۵,۴۵۵	۰.۴۷۰	۰.۴۹۴	۹	۰.۵۸۲	۰,۱۶۴

شاخص‌کل- هفتگی	Correlation- Attribute Ranking	۸۱.۶۳۶	۰,۲۰۴	۰,۳۸۳	۱۰	۰,۸۶۶	۰.۶۲۲
شاخص‌کل هم وزن- روزانه	Symmetrical- Attribute Ranking	۶۵.۱۹۳	۰.۴۱۳	۰.۴۸۵	۱۲	۰.۶۸۱	۰,۲۹۵
شاخص‌کل هم وزن- هفتگی	Symmetrical- Attribute Ranking	۸۸,۲۱۴	۰,۱۴۴	۰,۳۳۴	۱۸	۰,۸۷۸	۰,۷۶۳
شاخص‌کل- روزانه	Symmetrical- Attribute Ranking	۵۷.۴۵۵	۰.۴۶۶	۰.۵۰۷	۱۰	۰.۵۶۶	۰,۰۸۷
شاخص‌کل- هفتگی	Symmetrical- Attribute Ranking	۸۵.۲۷۳	۰,۱۷۳	۰,۳۶۹	۹	۰,۸۶۹	۰.۶۹۶
شاخص‌کل هم وزن- روزانه	No Filter	۶۳,۱۶۸	۰.۴۰۰	۰.۵۰۲	۰	۰.۶۷۵	۰,۲۵۳
شاخص‌کل هم وزن- هفتگی	No Filter	۸۷,۸۴۵	۰,۱۴۰	۰,۳۳۹	۰	۰,۸۷۸	۰,۷۵۶
شاخص‌کل- روزانه	No Filter	۵۸,۰۰۰	۰.۴۵۹	۰.۵۲۳	۰	۰.۵۸۲	۰,۱۱۶
شاخص‌کل- هفتگی	No Filter	۸۴.۳۶۴	۰,۱۷۰	۰,۳۸۴	۰	۰,۸۵۹	۰.۶۸۰

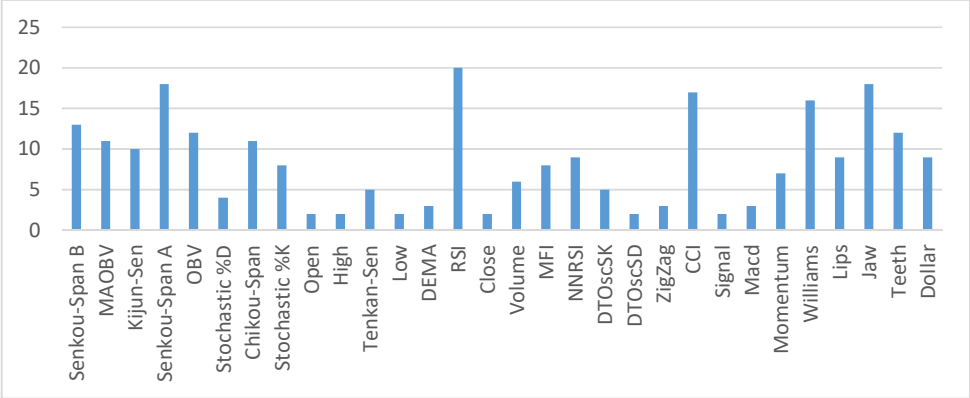
(منبع: یافته‌های پژوهشگر)

جدول ۳: نتایج الگوریتم‌های انتخاب ویژگی و دسته‌بندی ماشین بردار پشتیبان

نماد- دوره زمانی	روش جستجو- مدل ارزیابی	ACC	MAE	RMSE	تعداد ویژگی‌های منتخب	AUC-ROC	MCC
شاخص‌کل هم وزن- روزانه	CFS- BEST FIRST	۶۶,۶۶۷	۰,۳۳۳	۰.۵۷۷	۱۳	۰.۶۶۲	۰,۳۲۷
شاخص‌کل هم وزن- هفتگی	CFS- BEST FIRST	۷۸,۲۶۹	۰,۲۱۷	۰.۴۶۶	۱۲	۰,۷۸۵	۰.۵۶۹
شاخص‌کل- روزانه	CFS- BEST FIRST	۵۹,۰۹۱	۰.۴۰۹	۰.۶۴۰	۵	۰.۵۸۱	۰,۱۶۴
شاخص‌کل- هفتگی	CFS- BEST FIRST	۶۵,۴۵۵	۰,۳۴۶	۰.۵۸۸	۵	۰.۶۴۰	۰,۲۸۲
شاخص‌کل هم وزن- روزانه	InfoGain- Attribute Ranking	۶۷.۴۰۳	۰,۳۲۶	۰.۵۷۱	۱۳	۰.۶۷۰	۰,۳۴۲
شاخص‌کل هم وزن- هفتگی	InfoGain- Attribute Ranking	۷۸,۲۶۹	۰,۲۱۷	۰.۴۶۶	۲۱	۰,۷۸۵	۰.۵۶۹
شاخص‌کل- روزانه	InfoGain- Attribute Ranking	۵۷.۶۳۶	۰.۴۲۴	۰.۶۵۱	۷	۰.۵۶۵	۰,۱۳۱
شاخص‌کل- هفتگی	InfoGain- Attribute Ranking	۷۴.۳۶۴	۰,۲۵۶	۰.۵۰۶	۱۱	۰,۷۵۵	۰.۵۰۴
شاخص‌کل هم وزن- روزانه	GainRatio- Attribute Ranking	۶۶,۴۸۳	۰,۳۳۵	۰.۵۷۹	۱۰	۰.۶۵۹	۰,۳۲۲
شاخص‌کل هم وزن- هفتگی	GainRatio- Attribute Ranking	۷۸,۲۶۹	۰,۲۱۷	۰.۴۶۶	۳۰	۰,۷۸۵	۰.۵۶۹
شاخص‌کل- روزانه	GainRatio- Attribute Ranking	۶۰,۳۶۴	۰,۳۹۶	۰.۶۳۰	۳۰	۰.۵۹۷	۰,۱۹۴
شاخص‌کل- هفتگی	GainRatio- Attribute Ranking	۷۳.۴۵۵	۰,۲۶۶	۰.۵۱۵	۱۰	۰,۷۳۸	۰.۴۷۰
شاخص‌کل هم وزن- روزانه	Correlation- Attribute Ranking	۶۷,۹۵۶	۰,۳۲۰	۰.۵۶۶	۸	۰.۶۷۵	۰,۳۵۳
شاخص‌کل هم وزن- هفتگی	Correlation- Attribute Ranking	۷۲,۱۹۲	۰,۲۷۸	۰.۵۲۷	۱۰	۰,۷۱۸	۰.۴۴۰
شاخص‌کل- روزانه	Correlation- Attribute Ranking	۵۷.۶۳۶	۰.۴۲۴	۰.۶۵۱	۹	۰.۵۶۵	۰,۱۳۲
شاخص‌کل- هفتگی	Correlation- Attribute Ranking	۶۵,۴۵۵	۰,۳۴۶	۰.۵۸۸	۱۰	۰.۶۴۰	۰,۲۸۲
شاخص‌کل هم وزن- روزانه	Symmetrical- Attribute Ranking	۶۷.۴۰۳	۰,۳۲۶	۰.۵۷۱	۱۲	۰.۶۷۰	۰,۳۴۲
شاخص‌کل هم وزن- هفتگی	Symmetrical- Attribute Ranking	۷۸,۰۸۵	۰,۲۱۹	۰.۴۶۸	۱۸	۰,۷۸۳	۰.۵۶۵
شاخص‌کل- روزانه	Symmetrical- Attribute Ranking	۵۸.۵۴۶	۰.۴۱۵	۰.۶۴۴	۱۰	۰.۵۷۵	۰,۱۵۱

شاخص کل- هفتگی	Symmetrical- Attribute Ranking	۷۳,۰۹۱	۰,۲۶۹	۰.۵۱۹	۹	۰,۷۴۰	۰.۴۷۳
شاخص کل هم وزن- روزانه	No Filter	۷۰,۹۰۲	۰,۲۹۱	۰.۵۳۹	۰	۰,۷۰۸	۰.۴۱۵
شاخص کل هم وزن- هفتگی	No Filter	۷۸,۲۶۹	۰,۲۱۷	۰.۴۶۶	۰	۰,۷۸۵	۰.۵۶۹
شاخص کل- روزانه	No Filter	۶۰,۳۶۴	۰,۳۹۶	۰.۶۳۰	۰	۰.۵۹۷	۰,۱۹۴
شاخص کل- هفتگی	No Filter	۷۳.۴۵۵	۰,۲۶۶	۰.۵۱۵	۰	۰,۷۴۵	۰.۴۸۳

(منبع: یافته‌های پژوهشگر)



شکل ۱: بسامد اندیکاتورهای انتخاب شده توسط الگوریتم‌های انتخاب ویژگی (منبع: یافته‌های پژوهشگر)

جدول ۴: نتایج هشت اندیکاتور بیشتر انتخاب شده با دسته‌بندی جنگل تصادفی

نماد- دوره زمانی	ACC	MAE	RMSE	AUC-ROC	MCC
شاخص کل هم وزن-روزانه	۶۹,۶۱۳	۰,۳۶۹	۰,۴۵۷	۰,۷۴۹	۰,۳۸۹
شاخص کل هم وزن- هفتگی	۹۰,۴۲۳	۰,۱۵۷	۰,۲۶۱	۰,۹۶۷	۰,۸۰۷
شاخص کل- روزانه	۵۸	۰,۴۴۱	۰,۵۰۴	۰,۶۲۶	۰,۱۵۴
شاخص کل- هفتگی	۸۸,۳۶۳	۰,۱۹	۰,۲۸۳	۰,۹۶۱	۰,۷۵۹

(منبع: یافته‌های پژوهشگر)

جدول ۵: مقایسه نتایج ۳۰ ویژگی و هشت اندیکاتور منتخب با دسته‌بندی جنگل تصادفی برای نمادهای فاسمین و بپاس

نماد- دوره زمانی	ACC	MAE	RMSE	AUC-ROC	MCC
فاسمین- روزانه- ۸ اندیکاتور منتخب	۶۲,۷۲۶	۰,۴۳	۰,۴۸۷	۰,۶۶۸	۰,۸۱۹
فاسمین- روزانه- ۳۰ ویژگی	۶۲,۳۹۷	۰,۴۳۷	۰,۴۸۲	۰,۶۷۰	۰,۸۲۶
بپاس- روزانه- ۸ اندیکاتور منتخب	۶۶,۶۱۲	۰,۳۸۸	۰,۴۶۹	۰,۶۸۷	۰,۷۶۷

بپاس- روزانه- ۳۰ ویژگی	۶۶,۲۸۱	۰,۳۸۸	۰,۴۵۸	۰,۷۰۳	۰,۷۸۶
فاسمین- هفتگی- ۸ اندیکاتور منتخب	۸۴,۵۶۵	۰,۲۵۳	۰,۳۳۷	۰,۹۲۰	۰,۳۶۶
فاسمین- هفتگی- ۳۰ ویژگی	۸۳,۴۱۵	۰,۲۷۹	۰,۳۴۱	۰,۹۲۸	۰,۳۶۷
بپاس- هفتگی- ۸ اندیکاتور منتخب	۸۸,۰۹۹	۰,۲۰۷	۰,۳۰۶	۰,۹۳۱	۰,۱۷۰
بپاس- هفتگی- ۳۰ ویژگی	۸۷,۲۷۳	۰,۲۲۸	۰,۳۲۰	۰,۹۱۵	۰,۱۷۱

(منبع: یافته‌های پژوهشگر)

نتیجه‌گیری و بحث

جهت کاهش ریسک سرمایه‌گذاری و افزایش میزان سود، پیش‌بینی روند تغییرات شاخص بورس برای اکثر سرمایه‌گذاران بسیار با اهمیت است، لذا در این پژوهش از روش‌های انتخاب ویژگی و دسته‌بندی جهت پیش‌بینی روند تغییرات شاخص‌کل و شاخص‌کل هم‌وزن در بازار بورس تهران با استفاده از ۳۰ ویژگی و اندیکاتور مختلف استفاده شد. پیش‌بینی‌ها در بازه زمانی روزانه و هفتگی انجام گردید. نتایج تجربی نشان می‌دهد که روش دسته‌بندی جنگل تصادفی در بسیاری از موارد از دقت بیشتری نسبت به سایر روش‌ها برخوردار است و با کاهش تعداد ویژگی‌ها و اندیکاتورها، می‌توان به پیش‌بینی دقیق‌تری (به‌ویژه در پیش‌بینی‌های دوره زمانی هفتگی) و در عین حال کاهش عملیات پردازشی دست یافت. علاوه بر این، با توجه به نتایج مشخص است که دقت پیش‌بینی روند تغییرات شاخص‌کل و شاخص‌کل هم‌وزن در دوره زمانی هفتگی به طور قابل توجهی بیشتر از دقت پیش‌بینی در دوره زمانی روزانه است، زیرا در بازه زمانی روزانه معمولا به علت وجود نوسان‌های کوتاه مدت ناشی از اخبار روزانه، نویزهای فراوانی وجود دارد که این نویزها سبب می‌شود تحلیل تکنیکال با مشکل مواجه شود و به همین علت تحلیل تکنیکال در بازه‌های زمانی کوتاه مدت نتایج قابل قبولی ندارد. همچنین رفتار گله‌ای[۲۴] در معاملات روزانه بسیار بیشتر از معاملات هفتگی یا ماهانه رایج است، که در آن بسیاری از معامله‌گران اطلاعات و تحلیل‌های شخصی خود را نادیده می‌گیرند و به تقلید از معامله‌گرانی می‌پردازند که آنها را مطلع می‌پندارند و به حرکات گله‌ای وارد می‌شوند. اخبار جعلی و رفتار گله‌ای معامله‌گران ناآگاه ایجاد نوسانات قیمت می‌کنند، اما تأثیر این نوسانات بسیار کوتاه می‌باشد طوری که در بازه زمانی هفتگی بسیار کمتر دیده می‌شود. با توجه به نتایج تجربی، اندیکاتورهای RSI، Jaw، Senkou-Span A، CCI، Williams %R، Senkou-Span B، Teeth و OBV نسبت به سایر اندیکاتورها و ویژگی‌ها بیشتر توسط الگوریتم‌های انتخاب ویژگی انتخاب شده‌اند، بنابراین احتمالا می‌توانند در تشخیص روند شاخص‌کل و شاخص‌کل هم‌وزن نسبت به سایر ویژگی‌ها و اندیکاتورها موثرتر باشند. با توجه به نتایج دسته‌بندی با اندیکاتورهای مذکور، تنها با محاسبه این هشت اندیکاتور می‌توان روند تغییرات شاخص‌ها را با دقت مناسب پیش‌بینی کرد. در

سیستم معالاتی ایچیموکو، Senkou-Span A و Senkou-Span B تشکیل دهنده ابر کومو می‌باشند. ابر کومو از جمله اندیکاتورهای معدودی است که دیدی رو به آینده دارد و معامله‌گران با استفاده از آن قادر خواهند بود بهترین سیگنال‌های ورود و خروج از بازار را دریافت کنند. از اینرو ابر کومو در تحلیل تکنیکال از اهمیت ویژه‌ای برخوردار است. Jaw و Teeth جزو سیستم معاملاتی تمساح[۲۵] هستند. فک تمساح[۲۶] میانگین متحرک ۱۳ روزه است که بر روی ۸ کندل آخر اعمال شده است و دندان تمساح[۲۷] میانگین متحرک ۸ روزه است که بر روی ۵ کندل آخر اعمال شده است. این دو اندیکاتور به شناسایی شروع و پایان یک روند در بازه زمانی مشخص و شناسایی روندهای قوی در یک سهم کمک می‌کنند. اندیکاتور Williams %R برای شناسایی نواحی اشباع خرید و اشباع فروش استفاده می‌شود. این اندیکاتور به معامله‌گر می‌گوید که قیمت فعلی نسبت به بالاترین قیمت در ۱۴ دوره گذشته کجا قرار دارد. اندیکاتور RSI که یکی از محبوب‌ترین اندیکاتورهای تکنیکال است، برای تشخیص قدرت روند و جهت آن استفاده می‌شود. معمولا از RSI برای استراتژی‌های معاملاتی خرید در پایین‌ترین قیمت[۲۸] و فروش در بالاترین قیمت[۲۹] استفاده می‌شود. در RSI دو ناحیه اشباع خرید و اشباع فروش وجود دارد که نشان دهنده کاهش خرید و فروشهای افراطی می‌باشد. اندیکاتور OBV یا حجم تعادلی[۳۰] تغییرات حجم را بر اساس حرکت قیمتی سهم می‌سنجد تا بتواند تغییرات قیمتی آینده را پیش‌بینی کند. از آنجا که تغییرات حجم در معاملات از اهمیت بالایی برخوردار است، استفاده از این اندیکاتور می‌تواند بااهمیت باشد، زیرا حجم می‌تواند نیروی محرک بازار باشد و تغییرات آن بسیار اهمیت دارد. اندیکاتور CCI یا شاخص کانال کالا[۳۱] شتاب نوسانات و میزان هیجانات موجود در سهم را نشان می‌دهد، این اندیکاتور قیمت فعلی سهم را با یک میانگین قیمتی مقایسه می‌کند و افزایش مقدار اندیکاتور به معنی افزایش هیجان در خریداران و کاهش مقدار و منفی شدن آن به معنی افزایش هیجان در فروشندگان است. نتایج دسته‌بندی دو نماد فاسمین و بپاس که در جدول ۵ آمده نیز بیانگر این است که هشت اندیکاتور منتخب برای این دو نماد هم با دقت مناسبی دسته‌بندی شده‌اند و حتی در مورد این دو نماد، دقت دسته‌بندی با هشت اندیکاتور منتخب بیشتر از دسته‌بندی با ۳۰ اندیکاتور اولیه است.

پیشنهاد می‌شود که در پژوهش‌های آینده از سایر الگوریتم‌های انتخاب ویژگی و دسته‌بندی و همچنین از الگوریتم‌های رگرسیون نیز استفاده شود، و همچنین داده‌های سایر بازارها، مانند بازار طلا و ارزهای دیجیتال، مورد بررسی قرار گیرد.

منابع

1) Alexander, S. S. (1961), Price movements in speculative markets: trends or random walks, Industrial Management Review 2, 7–26.

2) Alexander, S. S. (1964), Price movements in speculative markets: trends or random walks, Number 2, Industrial Management Review 5, 25–46.

3) Biau, G., & Scornet, E. (2016). A random forest guided tour. Test, 25(2), 197-227.

4) Boobalan, C. (2014). Technical analysis in select stocks of Indian companies. *International Journal of Business and Administration Research Review*, *2*(4), 26-36.

5) Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2016). WEKA manual for version 3-8-1. University of Waikato, Hamilton, New Zealand.

6) Chandar, S. K. (2020). Grey Wolf optimization-Elman neural network model for stock price prediction. Soft Computing, 1-10.

7) Chen, L., Qiao, Z., Wang, M., Wang, C., Du, R., & Stanley, H. E. (2018). Which artificial intelligence algorithm better predicts the Chinese stock market?. IEEE Access, 6, 48625-48633.

8) Chong, T. T. L., & Ng, W. K. (2008). Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30. Applied Economics Letters, 15(14), 1111-1114.

9) Chung H, Shin KS. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. Sustainability 10:1–18.

10) Dash, M. and Liu, H. (1997). Feature selection for classification. Intelligent data analysis 1:131-156.

11) Dash R, Dash PK. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. The Journal of Finance and Data Science. 2(1), 42–57.

12) Fama, E. F. (1965), The behavior of stock market prices, Journal of Business 38, 34–105.

13) Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. Machine learning, 31(1), 1-8.

14) Guo ZQ, Wang HQ, Liu Q (2013) Financial time series forecasting using LPP and SVM optimized by PSO. Soft Comput 7(5):805–818.

15) Lei L (2018) Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. Appl Soft Comput 62:923–932.

16) Liao Z, Wang J (2010) Forecasting model of global stock index by stochastic time effective neural networks. Expert Syst Appl 37(1):834–841.

17) Lin, X., Yang, Z., Song, Y. (2009). Short-term stock price prediction based on echo state networks. Expert Syst. Appl. 36(3), 7313–7317.

18) Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Structure, 405(2), 442-451.

19) Miner, R. C. (2008). High probability trading strategies: Entry to exit tactics for the forex, futures, and stock markets (Vol. 328). John Wiley & Sons.

20) Murphy JJ. (1999). Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance Series. New York Institute of Finance.

21) Nahil, A, Lyhyaoui, A. (2018). Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of Casablanca stock exchange. Procedia Comput. Sci. 127, 161–169.

22) Patel J, Shah S, Thakkar P, Kotecha K. (2015). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications. 42(1), 259–268.

23) Rahimunnisa K (2019) Hybrdized genetic-simulated annealing algorithm for performance optimization in wireless adhoc network. J Soft Comput Paradigm (JSCP) 1(01):1–13

24) Ray P, Mahapatra GS, Rani P, Pandy SK, Dey KN (2014) Robust feed forward and recurrent neural Network based dynamic weighted combination models for software reliability prediction. Appl Soft Comput 22:629–637.

25) Talebi, H., Hoang, W., & Gavrilova, M. L. (2014). Multi-scale foreign exchange rates ensemble for classification of trends in forex market. Procedia Computer Science, 29, 2065-2075.

26) Witten, I.H., Frank, E. and Hall, M.A. (2011) Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edition, Morgan Kaufmann Publishers, Burlington.

27) Wong, H. B., & Lim, G. H. (2011). Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV. Proceedings of Singapore healthcare, 20(4), 316-318.

28) Xie, H., Fan, K., & Wang, S. (2021). Candlestick Forecasting for Investments: Applications, Models and Properties. Routledge.

29) Zhong, X., Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction. Expert Syst. Appl. 67, 126–139.

یادداشت‌ها:

1 BPNN (Back Propagation Neural Network)
2 FLANN (Functional Link Artificial Neural Network)
3 WNN (Wavelet Neural Network)
4 RNN (Recurrent Neural Network)
5 RBNN (Radially Bounded Nearest Neighbor)
6 GWO-ENN (Grey Wolf Optimization- Elman Neural Network)
7 KPCA (Kernel Principal Component Analysis)
8 SVM (Support Vector Machine)
9 Naive Bayes
10 Fast Robust Principal Component Analysis
11 Back Propagation
12 ELM
13 RBF
14 Search Method
15 Attribute Evaluator
16 CFS (Correlation Based Features Selection Subset Evaluator)
17 Correlation Attribute Evaluator
18 InfoGain (Information Gain Attribute Evaluator)
19 GainRatio (Gain Ratio Feature Evaluator)
20 Symmetrical Attribute Evaluator
21 Random Forest
22 Support Vector Machine
23 Matthews Correlation Coefficient
24 Herding Behavior
25 Alligator
26 Gator Jaw
27 Gator Teeth
28 Over Bought
29 Over Sold
30 On Balance Volume
31 Commodity Chanel Index

Identification of effective indicators on predicting trends of total index of Tehran Stock Exchange using feature selection and classification algorithms

Mohammad Javad Sheikhzadeh[1]
Sajjad Rahmani[2]

Receipt: 09/06/2022 **Acceptance: 12/09/2022**

Abstract

Because of the numerous environmental, industrial, micro, and macro elements that influence the index and stock price trend, it is undeniably difficult to predict changes in the index and stock price trend. Although the aforementioned factors are difficult or impossible to measure in some cases, micro factors such as price history and trade volume are simply measurable and available. The goal of this research is to use feature selection and classification algorithms to find the most effective features and indicators for predicting the total index and total weighted index. Then we will examine the proposed model in the daily and weekly time period from January 2020 to May 2022. The results show that it is possible to predict the trend of changes with high accuracy using a limited number of indicators, and that the prediction accuracy is much higher in the weekly time interval than in the daily time interval.

Keywords

Total Index, Indicator, Feature selection, Classification

1-Department of Computer Science, Faculty of Mathematics and Computer Science, Damghan University, Damghan, Iran. Sheikhzadeh@gonbad.ac.ir
2-Department of Computer Science, Faculty of Mathematics and Computer Science, Damghan University, Damghan, Iran. (Corresponding Author) Rahmani@du.ac.ir