راهبرد مدیریت مالی
سال چهارم، شماره چهاردهم
پاییز ۱۳۹۵
صص ۱۴۸-۱۲۱

دانشگاه الزهرا (س)
دانشکده علوم اجتماعی و اقتصادی
تاریخ دریافت: ۱۳۹۵/۰۴/۱۵
تاریخ تصویب: ۱۳۹۵/۰۶/۰۶

پیش‌بینی روند تغییرات قیمت سهم با استفاده از ماشین بردار پشتیبان وزن‌دهی شده و انتخاب ویژگی هیبرید به منظور ارائه استراتژی معاملاتی بهینه[۱]

سعید باجلان[۲]، سعید فلاحپور[۳] و ناهید دانا[٤]

چکیده

در این پژوهش، یک مدل پیش‌بینی براساس روش ماشین بردار پشتیبان تعدیل‌شده با استفاده از وزن‌دارکردن تابع جریمه مدل با توجه به حجم معاملات واقعی روزانه به منظور افزایش دقت پیش‌بینی نوسان‌های کوتاه مدت در بازار سهام و دست‌یابی به استراتژی معاملاتی بهینه، ارائه شده است. همراه با طبقه‌بندی‌کننده ماشین بردار پشتیبان تعدیل‌شده، از یک روش انتخاب ویژگی هیبرید، مرکب از یک بخش فیلترکننده و یک بخش پوشش‌دهنده به منظور انتخاب زیرمجموعه‌ای بهینه از ویژگی‌ها استفاده شده است. همچنین به منظور بررسی توانایی مدل پیشنهادی در پیش‌بینی روند قیمت، یک استراتژی معاملاتی بر پایه نتایج مدل داده می‌شود. ورودی مدل چندین شاخص تحلیل تکنیکال و شاخص‌های آماری متعددی هستند که برای تعداد ۱۰ سهم انتخاب شده از بورس اوراق بهادار تهران محاسبه شده‌اند. نتایج نشان می‌دهد که مدل ماشین بردار پشتیبان وزن‌دهی شده، همراه با روش انتخاب ویژگی هیبرید پیشنهاد شده، میزان دقت پیش‌بینی را به میزان قابل توجهی افزایش داده و نیز نتایج استراتژی معاملاتی پیشنهادشده را نسبت به استراتژی‌های رقیب، هم از لحاظ میزان بازده کلی و هم از لحاظ میزان بیشینه ضرر در طول دوره سرمایه‌گذاری بهبود می‌بخشد.

واژه های کلیدی: ماشین بردار پشتیبان، انتخاب ویژگی، پیش‌بینی رون، استراتژی معاملاتی
طبقه‌بندی موضوعی: C38، C53، G11، G12، G17

---

۱. کد DOI مقاله: 10.22051/jfm.2016.2575
۲. استادیارگروه مالی و بیمه، دانشکده مدیریت دانشگاه تهران، saeedbajalan@ut.ac.ir
۳. استادیارگروه مالی و بیمه، دانشکده مدیریت دانشگاه تهران، Falahpor@ut.ac.ir
٤. دانشجوی کارشناسی ارشد رشته مهندسی مالی، دانشگاه تهران، نویسنده مسئول، nahiddana@gmail.com

مقدمه

پیش‌بینی دقیق و قابل اطمینان روندهای قیمتی در بازار سهام به عنوان یکی از تاثیرگذارترین عوامل بر سوددهی معاملات فعال، همواره حوزه‌ای مورد توجه و بحث در موضوع‌های مالی جدید بوده و هست. پیش‌بینی روند قیمت سهام به علت ماهیت نامانا و پرنوسان بازار قیمت سهام کاری پیچیده است و به طور معمول نتایج حاصل از آن همواره با میزانی از عدم اطمینان همراه است. بنابراین کارهای زیادی در این حوزه انجام و روش‌ها و تکنیک‌های متعددی داده شده است که می‌توان آن‌ها را در دو دسته کلی طبقه‌بندی کرد: گروه نخست، روش‌ها و تکنیک های مبتنی بر این فرض هستند که قیمت یک دارایی، روند و سطوح آن در آینده را می‌توان با تجزیه و تحلیل عملکرد گذشته آن به دست آورد که این روش‌ها شامل مدل‌سازی‌های آماری و اقتصاد سنجی و همین‌طور روش‌های تحلیل تکنیکال می‌باشد. اما گروه دوم شامل روش‌هایی است که دارایی را از یک دید وسیع‌تر مثل اقتصاد محیط دارایی، احساس‌ها و منطق سرمایه‌گذاران و فاکتورهای دیگر مورد بررسی قرار می‌دهد که به راحتی قابل کمی‌سازی و اندازه‌گیری نمی‌باشند(زبیکوسکی ۲۰۱۴).

پیش‌بینی دقیق تغییرات قیمت سهام در دادن یک استراتژی معاملاتی کارآمد و سودبخش از اهمیت بسیاری برخوردار است (لیونگ، داک و چن، ۲۰۰۰). پژوهش‌های اخیر در زمینه پیش‌بینی روند تغییرات قیمت سهام حاکی از این است که روش‌های جدید داده‌کاوی و مدل‌های غیر خطی، عملکرد بهتری نسبت به روش‌های سنتی مانند لوجیت، پروبیت و گشت تصادفی دارند(فلاح‌پور، گل‌ارضی و فتوره‌چیان ۱۳۹۲).

طبق مطالعات صورت گرفته و تجربیات منتج از بازار سهام، روندهای قیمتی در بازار سهام تحت تاثیر حجم معاملات سهم هستند و دخالت دادن این مقدار در طبقه‌بندی‌کننده[1]، به میزان چشمگیری دقت پیش‌بینی تغییر جهت‌های روند قیمت را بهبود داده و به تولید سیگنال‌های خرید و فروش به موقع می‌انجامد(چاوارناکول،و انک ۲۰۰۸).

یکی از بهبودهای به کار رفته در مطالعات جدید، با توجه به پیشرفت‌های کاربرد روش‌های هوش مصنوعی، علاوه بر کاربرد مدل‌های هیبریدی، استفاده از انتخاب ویژگی به عنوان یک پیش مرحله برای مدل طبقه‌بندی کننده اصلی می‌باشد. هرچند طبق مطالعات انجام شده، ماشین بردار پشتیبان، عملکرد جالب توجه و مناسبی در پیش‌بینی و طبقه‌بندی دارد، اما دقت عملکرد آن به میزان

---

1. Classifier

قابل توجهی تحت تاثیر تعداد متغیرهای ویژگی ورودی آن است. بنابراین کاهش تعداد ویژگی‌هایی که باید در آموزش ماشین بردارپشتیبان به کارگرفته شود، تاثیر به سزایی در افزایش دقت نتایج و کاهش هزینه دارد(لی ۲۰۰۹، پینگ لی ۲۰۱۱ و ژانگ ۲۰۱۴).

پژوهش‌های موجود نشان داده‌اند که وابسته‌سازی مدل به حجم معاملات و به کارگیری تجزیه و تحلیل تکنیکال به صورت مجزا عملکرد مدل را بهبود قابل توجه می‌بخشند. بنابراین حساس‌سازی نتایج مدل ماشین بردارپشتیبان به حجم معاملات و استفاده از انتخاب ویژگی به منظور کاهش ابعاد و افزایش سرعت و دقت، می‌تواند سبب بهبودهای چشمگیری در نتایج شود.

مبانی نظری و مروری بر پیشینه پژوهش

تعداد متغیرهای لازم برای اندازه گیری هر مشاهده اندازه "ابعادداده "نامیده می‌شود. زمانی که داده‌ها دارای ابعاد زیاد هستند، علی‌رغم فرصت‌هایی که به وجود می‌آورند، چالش‌های محاسباتی زیادی ایجاد می‌کنند. یکی از مشکلات کار با داده‌های به ابعاد زیاد این است که تمامی ابعاد داده‌ها مرتبط با آن چیزی نیست که الگوریتم یادگیری سعی در استخراج آن دارد و وجود آن‌ها در داده‌ها، کارایی به کارگیری الگوریتم یادگیری را کاهش، هزینه محاسباتی را افزایش و ریسک "بیش‌برازش " را افزایش می‌دهد. انتخاب ویژگی با تعیین زیرمجموعه‌ای از ویژگی‌های موجود که بیشترین اهمیت را برای مساله طبقه‌بندی دارند، ابعاد مجموعه اولیه ویژگی‌ها را کاهش می‌دهد (لیو و ژنگ ۲۰۰۶).

روش‌های مختلف انتخاب ویژگی تلاش می‌کنند تا از میان 2^N زیر مجموعه ممکن، زیرمجموعه‌ای را انتخاب کنند که علاوه بر توصیف کارامد داده‌های ورودی، در عین حفظ دقت نتایج پیش‌بینی، متغیرهای نویزی و نامربوط را کاهش‌می‌دهد (گویان و الیسوف ۲۰۰۳). برخلاف روش‌های مبتنی بر استخراج ویژگی، این نوع روش‌ها معنای اصلی ویژگی‌ها را بعد از کاهش حفظ می‌کنند(فودور، ۲۰۰۲).

برای حذف یک ویژگی نامرتبط، باید یک معیار انتخاب ویژگی تعریف‌شود که بتواند میزان مرتبط بودن هر ویژگی را با نوع یا برچسب خروجی‌ها بسنجد. زمانی‌که معیار انتخاب ویژگی مشخص می‌شود، باید یک دستورالعمل نیز برای یافتن زیرمجموعه‌ای از ویژگی‌ها انتخاب‌شود.

چراکه با افزایش تعداد ویژگی‌ها، ارزیابی مستقیم تمام 2^N زیرمجموعه ممکن، به سمت تبدیل‌شدن به یک مساله با زمان حل نامعلوم[1] می‌شود (چاندراشکار ۲۰۱۴).

روش‌های حذف یا کاهش ویژگی به سه دسته روش‌های فیلتر[2]، پوششی[3] و محاطی[4] تقسیم‌بندی می‌شوند(کوهاوی ۱۹۹۷ و لانگلی ۱۹۹۴).

در ادامه، شرح مختصری بر هر یک از این رویکرد می‌آوریم.

روش‌های فیلترکننده از تکنیک‌های رتبه‌بندی ویژگی‌ها به عنوان معیار اصلی در انتخاب ویژگی استفاده می‌کنند. براساس پژوهش‌ها علت به کارگیری این روش‌ها، سادگی و موفقیت‌هایی است که روش‌های رتبه‌بندی تاکنون داشته‌اند (چاندراشکار۲۰۱۴). در این رویکرد از یک معیار رتبه‌بندی مناسب برای امتیازدهی به ویژگی‌ها استفاده می‌شود و با تعیین یک حد آستانه، ویژگی‌های با امتیاز زیر این حد حذف می‌شود. از آن‌جایی‌که روش‌های رتبه‌بندی پیش از مرحله طبقه‌بندی و به منظور فیلترکردن و حذف ویژگی‌های نامربوط و یا کمتر مربوط به کار می‌روند، به آن‌ها روش‌های فیلترکننده گفته می‌شود. یک ویژگی خاص، این خاصیت پایه‌ای را دارد که شامل اطلاعات مفیدی در مورد گروه‌های مختلف موجود در داده‌ها باشد. این خاصیت را می‌توان به عنوان میزان مرتبط بودن ویژگی تعریف نمود که شاخصی از میزان سودمند بودن ویژگی در طبقه‌بندی گروه‌های مختلف داده است(کوهاوی ۱۹۹۷).

روش انتخاب ویژگی فیشر که در این پژوهش به عنوان بخشی از فرآیند انتخاب ویژگی هیبرید پیشنهاد شده است، یک روش فیلترکننده ساده اما پایه‌ای می‌باشد که میزان تمایز و تفاوت بین دو مجموعه داده را می‌سنجد (چن و لین ۲۰۰۵). روش انتخاب ویژگی فیشر برای هر ویژگی در مجموعه اولیه از ویژگی‌ها، با توجه به تساوی زیر، یک امتیاز محاسبه می‌کند. با داشتن بردار داده‌های آموزش به صورت x_k ،k=1,2,…,m و n_+ و n_- به ترتیب، به عنوان تعداد نمونه‌های مثبت و تعداد نمونه‌های منفی امتیاز فیشر برای i امین ویژگی به صورت زیر محاسبه می‌شود:

$$F(i) \equiv \frac{\left(\bar{x}_i^{(+)} - \bar{x}_i\right)^2 + \left(\bar{x}_i^{(-)} - \bar{x}_i\right)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}\left(x_{k,i}^{(+)} - \bar{x}_i^{(+)}\right)^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}\left(x_{k,i}^{(-)} - \bar{x}_i^{(-)}\right)^2}$$

---

1. Nondeterministic polynomial time
2. Filter
3. Wrapper
4. Embeded

که در این عبارت، $\bar{x}_i$، $\bar{x}_i^{(+)}$ و $\bar{x}_i^{(-)}$ به ترتیب میانگین کل و میانگین ویژگی‌های مثبت و منفی می‌باشد.

$x_{k,i}^{(+)}$ و $x_{k,i}^{(-)}$ ویژگی iام از k امین ویژگی به ترتیب مثبت و منفی می‌باشد.

صورت کسر در تساوی بالا، تمایز بین مجموعه‌های مثبت و منفی را نشان می‌دهد و مخرج کسر، تفاوت‌ها را در داخل هر یک از دو مجموعه مثبت و منفی اندازه‌گیری می‌کند. هرچه میزان امتیاز فیشر محاسبه شده برای هر ویژگی بیشتر باشد، احتمال کمک بیشتر آن ویژگی به فرآیند طبقه‌بندی بکند، بیشتر است.

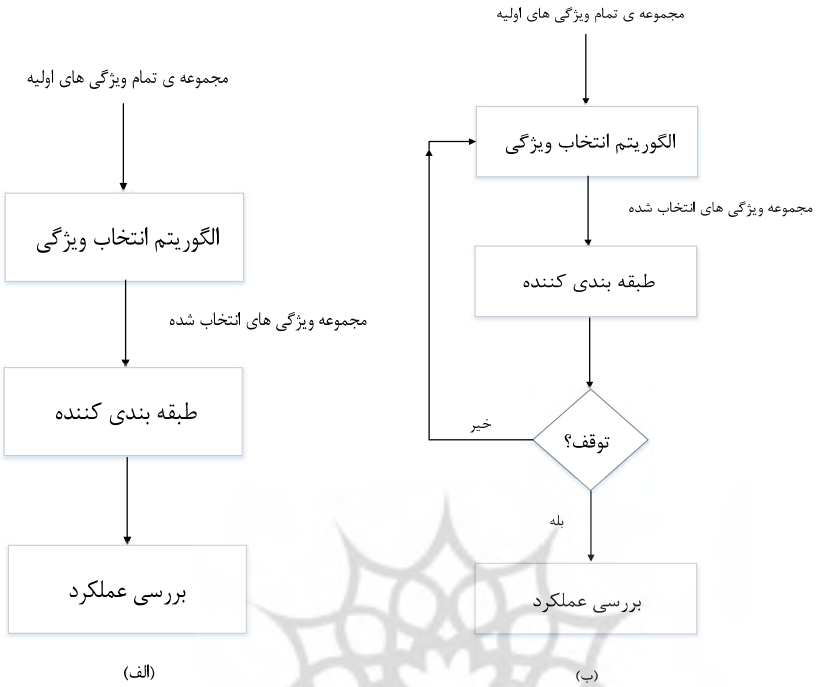
روش‌های پوششی از پیش‌بینی کننده به عنوان یک جعبه سیاه و از عملکرد پیش‌بینی کننده به عنوان تابع هدف برای ارزیابی زیرمجموعه انتخاب شده از ویژگی‌ها استفاده می‌کنند. از آنجایی‌که ارزیابی و بررسی همه 2^N حالت ممکن، کار بسیار پیچیده‌ای است، زیرمجموعه‌های بهینه محلی، توسط الگوریتم‌های فراابتکاری یافت می‌شوند و مورد بررسی قرار می‌گیرند. از یک‌سری الگوریتم‌های جست و جو می‌توان برای یافتن زیرمجموعه‌ای از ویژگی‌ها استفاده کرد که تابع هدف به عنوان عملکرد طبقه‌بندی کننده را بیشینه می‌کند. روش‌های شاخه و کران از درخت ساختار به منظور ارزیابی زیرمجموعه‌های مختلف از یک مجموعه ویژگی استفاده می‌کند (کوهاوی ۱۹۹۷و نارندرا ۱۹۹۷). اما این روش جست و جو برای زمان با تعداد ویژگی‌های خیلی زیاد، بسیار طولانی و پیچیده می‌شود(سعیز، ۲۰۰۷). بنابراین می‌توان از الگوریتم‌های ساده شده‌ای مثل جست و جوی ترتیبی و یا الگوریتم‌های تکاملی مانند الگوریتم ژنتیک (GA) و یا الگوریتم بهینه‌سازی ذرات (PSO) استفاده کرد که امکان دستیابی به بهینه محلی را فراهم می‌کنند، نتایج خوبی را می‌دهند و از نظر محاسباتی هم شدنی هستند(پودیل۱۹۹۴).

روش جست‌وجوی ترتیبی، شامل جست‌وجوی ترتیبی پیشرو[1] و جست‌وجوی ترتیبی پسرو[2] می‌باشد. در جست‌وجوی ترتیبی پیشرو که در این پژوهش به عنوان بخش دوم فرآیند انتخاب ویژگی هیبرید در نظرگرفته شده است، با یک زیر مجموعه خالی از ویژگی‌ها، F_0 شروع می‌کنیم و به ترتیب ویژگی‌ها را یک به یک از مجموعه اصلی به این زیر مجموعه اضافه می‌کنیم، تا زمانی که اندازه زیرمجموعه به تعداد مطلوب برسد.

در نمودار ۱، فرآیند پوششی نشان داده شده است.

---

1. Sequential forward search
2. Sequential backward search



نمودار ۱: (الف) نمای کلی مراحل انتخاب ویژگی در روش‌های فیلترکننده (ب) نمای کلی مراحل انتخاب ویژگی در روش‌های پوششی

هدف از رویکرد پوششی، کاهش زمان محاسباتی برای انتخاب زیرمجموعه ویژگی‌های انتخاب‌شده در فرآیند پوششی است (بلوم، ۱۹۹۷). این مدل از مزایای هردومدل قبلی به وسیله استفاده از معیارهای ارزیابی مختلفشان در مراحل متفاوت جست و جو سود می‌برد. در این روش وابستگی‌های ویژگی‌ها با هم در نظر گرفته می‌شود(گویان و وستون ، ۲۰۰۳). در متد محاطی، جستجو برای یک مجموعه ویژگی بهینه، درون ساختار دسته‌بندی کننده قرار دارد.

الگوریتم ماشین بردار پشتیبان برای طبقه‌بندی و دسته‌بندی، نخستین بار توسط فیشر در سال ۱۹۳۶ عرضه شد که تفاوت آن با سایر الگوریتم‌های یادگیری ماشینی در معیارآن برای بهینه‌سازی بود که کم کردن خطای طبقه‌بندی می‌باشد. در سال ۱۹۶۵، وپنیک نظریه یادگیری آماری را به نحوی مستحکم‌تر بیان نمود و طبقه‌بندی کننده‌های ماشین بردار پشتیبان را توسعه داد(فلاح‌پور، طبسی ۱۳۹۲).

ماشین بردار پشتیبان الگوریتمی است که به دنبال حداکثر کردن حاشیه ابر صفحه جدا کننده داده‌ها می‌باشد. حداکثر کردن حاشیه ابر صفحه به حداکثر شدن تفکیک بین طبقات می‌انجامد.

برای اندازه‌گیری این مقدار فاصله از نزدیکترین نقاط به ابرصفحه استفاده می‌کنند که این نقاط بردارهای پشتیبان نامیده می‌شوند(هوان و همکاران ۲۰۰۷ و شان و همکاران ، ۲۰۰۵).

طی سال‌های اخیر در حوزه‌های مختلف مالی مانند پیش‌بینی سری‌های زمانی و پیش‌بینی ورشکستگی از ماشین بردار پشتیبان استفاده‌های بسیاری شده است. در این پژوهش نیز ما با استفاده از ماشین بردار پشتیبان تعدیل‌شده با وزن‌دهی به پیش‌بینی روند قیمت سهام و ارائه استراتژی معاملاتی با استفاده از نتایج آن می‌پردازیم.

مبانی نظری و مروری بر پیشینهٔ پژوهش

آن چه که بر سودآوری یک استراتژی معاملاتی بیشترین تاثیر را دارد، پیش‌بینی به موقع روند قیمت‌ها می‌باشد. براساس فرضیه بازار کارای فاما (۱۹۷۰)، قیمت‌ها به سرعت نسبت به اطلاعات در دسترس واکنش نشان داده و با آن تغییر می کنند و تنها اطلاعات جدید تعیین کننده قیمت‌ها می‌باشد. بنا بر نظریه او، از آن‌جایی که اخبار قابل پیش‌بینی نمی‌باشد، برای حرکت قیمت سهام تصادفی و غیر قابل پیش‌بینی است. پس از فاما در چند دهه اخیر، پژوهش‌های زیادی در مخالفت با نظریه وی انجام شد(لاس، ۲۰۰۰ و تکسیرا و الیویرا، ۲۰۱۰). پژوهش‌ها در زمینه پیش‌بینی پذیر بودن قیمت دارایی‌های مالی را می‌توان در حوزه پژوهش‌هایی دانست که بر پایه سری‌های زمانی(وانگ ۲۰۰۵)، و طبقه‌بندی و کشف الگو(بااو و یانگ ۲۰۰۸)، بر پایه بهینه‌سازی(چانگو هسیو،۲۰۰۷) و همچنین بر پایه روش‌های هیبرید انجام شده‌اند(کیم و شین، ۲۰۰۷). با توجه به نتایج پژوهش‌های داخلی و خارجی بسیاری، برتری روش‌های ماشین بردار پشتیبان بر روش‌هایی از جمله روش شبکه‌های عصبی مصنوعی و نیز روش‌های آماری ثابت شده است(کریستیانی و همکاران، ۲۰۰۰ ، کیم ۲۰۰۳ و مین و لی ۲۰۰۵). بنابراین در ادامه به بیان پژوهش‌هایی می‌پردازیم که بیشترین ارتباط را با موضوع مورد بررسی ما دارند.

آبراهام و همکارانش (۲۰۰۱) با استفاده از شبکه عصبی انتشار پسرو[1]، جهت و روند حرکتی شاخص نزدک را پیش‌بینی کردند. آن‌ها از روش انتخاب ویژگی آنالیز مولفه اصلی[2] ، بر روی شاخص‌های تحلیل بنیادین استفاده کردند تا بیشترین دقت را به دست آورند. سه سال بعد از آن‌ها، به منظور مقایسه عملکرد مدل‌های پیش‌بینی کننده ماشین بردار پشتیبان و شبکه عصبی انتشار

1. Back-propagation neural network
2. Principal component analysis

پسرو(بریو و باتر، ۲۰۰۴)، اقدام به پیش‌بینی شاخص نزدک با استفاده از انتخاب ویژگی با روش آنالیز مولفه اصلی و آنالیز عاملی[1]، بر روی شاخص‌های تکنیکال کردند که نتایج، نشان‌دهنده برتری ماشین بردار پشتیبان بود.

چودریو گرگ (۲۰۰۸) با استفاده از ترکیب ماشین بردار پشتیبان و الگوریتم ژنتیک اقدام به پیش‌بینی روند حرکت قیمت سهام نمودند. نتایج پژوهش آن‌ها نشان داد که دقت مدل ترکیبی از مدل ماشین بردار پشتیبان به تنهایی بیشتر است.

هوانگ و همکاران (۲۰۰۸) به پیش‌بینی قیمت سهم با استفاده از روش انتخاب ویژگی پوششی همراه با یک سیستم مرکب از چندین طبقه‌بندی‌کننده شامل ماشین بردار پشتیبان و شبکه عصبی اقدام کردند. نتایج آن‌ها نشان داد که دقت پیش‌بینی‌ها با استفاده از روش انتخاب ویژگی پوششی نسبت به روش‌های فیلتر کننده مختلف بالاتر است.

لین و همکاران (۲۰۰۹) به پیش‌بینی شاخص S & p500 با استفاده از انتخاب ویژگی بر روی شاخص‌های تکنیکال و استفاده از مدل‌های پیش‌بینی کننده شبکه عصبی انتشار پسرو، شبکه انعکاس حالت[2] و شبکه عصبی بازگردنده[3] اقدام نمودند.

در پژوهشی بر روی شاخص بورس تایوان، لای و همکاران (۲۰۰۹)، از روش آماری رگرسیون و درخت تصمیم فازی برای پیش‌بینی جهت حرکت شاخص بورس تایوان استفاده شد. همچنین در این پژوهش از ۷ شاخص تکنیکال به عنوان ویژگی‌های ورودی مدل استفاده شد.

مینگ (۲۰۰۹) با هدف پیش‌بینی روند تغییرات قیمت شاخص نزدک، به انجام پژوهشی با ترکیب مدل ماشین بردار و انتخاب ویژگی هیبرید حاصل از ترکیب انتخاب ویژگی فیشر و جست و جوی ترتیبی پیشروی پشتیبانی شده اقدام نمود. وی به عنوان ویژگی‌های ورودی مدل از ۳۰ سهم موجود در این شاخص استفاده نمود. نتایج پژوهش مینگ نشان داد که استفاده از ماشین بردار پشتیبان همراه با انتخاب ویژگی هیبرید دقت بیشتری در پیش‌بینی روند نسبت به شبکه عصبی انتشار پسرو و نیز استفاده از سایر روش‌های انتخاب ویژگی دارد.

نیر و همکاران (۲۰۱۰)، از الگوریتم ژنتیک، بر پایه درخت تصمیم، همراه با پیش‌بینی با استفاده از ماشین بردار پشتیبان استفاده‌کردند. آن‌ها از شاخص‌های تکنیکال به عنوان ورودی‌های مدل

---

1. Factor analysis
2. Echo state network
3. Recurrent neural network

استفاده نمودند. نتایج این پژوهش نشان داد که مدل ترکیبی داده شده، از عملکرد بهتری نسبت به شبکه‌های عصبی مصنوعی برخوردار است.

در پژوهش دیگری منظور ارزیابی و مقایسه روش انتخاب ویژگی فرکتال[1] با سایر روش‌ها، از جمله رلیف[2]، سی اف اس[3] و چند روش‌دیگر، با استفاده از طبقه‌بندی کننده ماشین بردار پشتیبان توسط لی پینگ و همکاران (۲۰۱۱)، برتری روش فرکتال بر سایر روش‌های فیلتری انتخاب ویژگی اثبات شد.

ژانگ (۲۰۱۴) از یک روش انتخاب ویژگی تصادفی[4] همراه با ۷ مدل پایه‌ای پیش‌بینی همراه با ۱۸ ویژگی ورودی، برای پیش‌بینی روند تغییرات شاخص بورس شانگهای استفاده کرد. نتایج نشان داد که روش انتخاب ویژگی پیشنهادی ژانگ، از لحاظ دقت، عملکرد بسیار بهتری نسبت به سایر روش‌های انتخاب ویژگی پرکاربرد از جمله آنالیز مولف اصلی دارد.

در پژوهشی که توسط زیبکوسکی (۲۰۱۴) انجام شد، از ماشین بردار پشتیبان وزن‌دهی شده با حجم معاملات واقعی روزانه، همراه با انتخاب ویژگی فیشر، برای استراتژی معاملاتی بهینه استفاده شد. وی به عنوان ورودی از ۷ شاخص تحلیل تکنیکال به منظور پیش‌بینی روند قیمت سهام استفاده نمود. او با نشان دادن برتری عملکرد استراتژی معاملاتی برپایه مدل پیشنهادی خود، برتری این مدل بر مدل ماشین بردار پشتیبان ساده و ماشین بردار پشتیبان بدون انتخاب ویژگی را اثبات نمود.

در حوزه مطالعات پیش‌بینی روند، می‌توان به پژوهش فلاح‌پور، گل‌ارضی و فتوره‌چیان (۱۳۹۲) اشاره نمود. آن‌ها به پیش‌بینی روند حرکت و تغییرات قیمت سهام در بازار بورس اوراق بهادار تهران اقدام و متغیرهای ورودی ماشین بردار پشتیبان را توسط الگوریتم ژنتیک بهینه‌سازی نمودند. نتایج پژوهش آن‌ها نشان داد که ماشین بردار پشتیبان برپایه الگوریتم ژنتیک، دقت بسیار بیشتری در پیش-بینی نسبت به ماشین بردار پشتیبان ساده دارد.

به طور کلی از پژوهش‌های داخلی در مورد پیش‌بینی قیمت و شاخص، می‌توان به موارد زیر اشاره کرد.

سینایی، مرتضوی و تیموری (۱۳۸۴)، با پیش‌بینی شاخص بورس اوراق بهادار با استفاده از شبکه‌های عصبی نشان دادند که شبکه عصبی، عملکرد بهتری نسبت به مدل پیش‌بینی خطی ARIMA دارد.

---

1. Fractal
2. Relief
3. CFS
4. Causal feature selection

در پژوهشی که توسط منجمی، ابرازی و رعیتی (۱۳۸۸) انجام شد، قیمت سهام در بازار بورس اوراق بهادار، توسط شبکه عصبی فازی و الگوریتم ژنتیک پیش‌بینی شد. نتایج پژوهش آن‌ها نشان داد که مدل ترکیبی پیشنهادی هم از لحاظ دقت و هم از لحاظ سرعت عملکرد بهتری دارد.

در پژوهش عبادی (۱۳۸۸) برای پیش‌بینی شاخص قیمت سهام بازار بورس تهران با استفاده از شبکه‌های عصبی مصنوعی مشخص شد که شبکه عصبی عملکرد خوبی در پیش‌بینی این شاخص دارد. همچنین در پژوهش دیگری به منظور مطالعه تاثیر فاکتورهای رفتاری بر قیمت سهام توسط هاشمی (۱۳۸۹) بر روی سهام ۱۰ شرکت از شرکت‌های شاخص داوجونز مشخص شد که فاکتورهای رفتاری در قیمت ۹ سهم از ۱۰ سهم موثر بوده و مد نظر قرار دادن آن‌ها به افزایش دقت پیش‌بینی می‌انجامد.

فرضیه‌های پژوهش

با توجه به نتایج پژوهش‌های انجام شده، فرضیه پژوهش حاضر به صورت زیر بیان می‌شود:

نتایج استراتژی معاملاتی پیشنهادی با استفاده از روش انتخاب ویژگی هیبرید و طبقه‌بندی کننده تعدیل‌شده ماشین بردارپشتیبان، هم از لحاظ بازده کلی و هم بیشینه ضرر ممکن[1] در طول دوره معاملاتی، بهتر از سایر روش‌های رقیب می‌باشد.

روش‌شناسی پژوهش

در این پژوهش سعی شده است با استفاد از شیوه‌ای که در ادامه می‌آید، استراتژی معاملاتی بهینه‌ای به منظور معامله فعال سهام عرضه شود. ابتدا به توضیح طبقه‌بندی‌کننده ماشین بردار پشتیبان وزن‌دهی‌شده با حجم معاملات[2] واقعی روزانه و سپس روش انتخاب ویژگی هیبرید پیشنهاد شده پرداخته می‌شود.

جامعه آماری پژوهش شامل کلیه شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران می‌باشد. کیفیت بهتر داده‌های قیمتی مرتبط با این شرکت‌ها و تعداد اطلاعات بیشتر و دسترسی بهتر به اطلاعات آنها از جمله دلایل انتخاب ابن جامعه آماری می‌باشد. تعداد نمونه مورد استفاده ۱۰ شرکت، از شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌باشد که با نمونه‌گیری تصادفی انتخاب می‌شوند. این شرکت‌ها باید ویژگی‌های زیر را داشته باشند:

۱. جزء شرکت‌های سرمایه‌گذاری و بانک‌ها نباشند.

---

1. Maximum drawdown
2. Volume weighted support vector machine

۲. حداقل در ۶۰٪ روزهای معاملاتی، سهام آن‌ها معامله شده باشد(نماد باز بوده باشد).

۳. قبل از سال ۱۳۸۷ در بورس پذیرفته شده باشند.

داده‌های مربوط به سهام با این ویژگی‌ها برای بازه ۱۳۸۴/۰۸/۰۱ تا ۱۳۹۴/۰۸/۳۰ گردآوری شده است. در این پژوهش از نرم‌افزار مطلب برای انجام پیش‌بینی استفاده شده است. به طوریکه ابتدا با استفاده از داده‌های هر ۱۰ سهم، به اجرای انتخاب ویژگی هیبرید بر پایه ماشین بردار پشتیبان تعدیل شده پرداخته و زیرمجموعه بهینه از ویژگی‌ها را به دست‌آورده و سپس با این زیر مجموعه ماشین بردار پشتیبان تعدیل‌شده را برای هر سهم به صورت جداگانه اجراکرده و استراتژی معاملاتی مناسب هر سهم به دست آمده است. همین‌طور به منظور مقایسه نتایج حاصل از دو شاخص متوسط بازده سرمایه‌گذاری طی دوره و بیشینه ضرر رخ داده در طول دوره استفاده شده است.

هدف اصلی الگوریتم ماشین بردار پشتیبان، یافتن ابرصفحه جداکننده‌ای است که با حل مساله بهینه‌سازی زیر، حاشیه $||w||$ را بیشینه کند.

$$\min \frac{1}{2\left|\left|w\right|\right|^2} + \sum_{i=1}^{m} C_i \xi_i,$$

با محدودیت :

$$y_i(wx_i - b) \geq 1 - \xi_i$$
$$\xi_i \geq 0$$

که در آن ξ_i متغیر اسلک غیر صفر است و C یک پارامتر جریمه برای هر ورودی مشخص x_i و y_i مقدار هدف مربوط به آن است. در مدل ماشین بردار پشتیبان وزن‌دهی شده با حجم معاملات پیشنهادی ما، عبارت جریمه به صورت زیر تعریف می‌شود:

$$C_i = V_i C$$

که در آن

$$V_i = \frac{V_t}{\sum_{i=0}^{d} W_{t-i}}$$

که در آن V_t نشان‌دهنده حجم معامله شده از سهم در روز t ام و d طول پنجره زمانی معاملاتی است. با استفاده از ضرایب لاگرانژ α_i و μ_i می‌توانیم مساله بهینه‌سازی را به صورت زیر بازنویسی کنیم.

$$\max L(w, b, \alpha) = \frac{1}{2\|w\|^2} + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i(y_i(wx_i - b) - 1 + \xi_i) - \sum_{i=1}^{m}\mu_i\xi_i$$

با محاسبه مشتقات جزئی، می‌توانیم دوگان این مساله را به صورت زیر بنویسیم:

$$Q(\alpha) = \sum_{i=1}^{m}\alpha_i - 1/2\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \varphi(x_i)\varphi(x_j)$$

که این مساله باید با در نظرگرفتن محدودیت‌های زیر بیشینه شود.

$$\sum_{i=1}^{m}\alpha_i y_i = 0$$
$$\alpha_i \geq 0$$

در مجموع مراحل کار این پژوهش را از انتخاب ویژگی با روش پیشنهادی تا طبقه‌بندی با استفاده از مدل ماشین بردار پشتیبان وزن‌دهی شده، می‌توانیم به صورت زیر جمع‌بندی و خلاصه کنیم.

مرحله ۱: ابتدار مقدار امتیاز فیشر را برای هر ویژگی f_i محاسبه می‌کنیم.

مرحله ۲: امتیازهای محاسبه شده را به ترتیب نزولی و تعداد k ویژگی نخست را مرتب می‌کنیم که k عبارت از $K = [i * \frac{|F|}{4}]$ برای $\{1,2,\dots,m\} \in$ است. در این‌جا، |F|تعداد کل ویژگی‌ها می‌باشد و m بزرگترین عدد صحیحی می‌باشد که در رابطه $|F|/2^m \geq 1$ صدق کند.

مرحله ۳: ویژگی با بیشترین امتیاز را برای ساخت F_n^{av} انتخاب می‌کنیم.

مرحله۴: ویژگی بعدی را که باید به F_n^{av} اضافه کنیم، با توجه به تساوی $F_{n+1}^i = F_n \cup \{f_i\}$ ، برای $f_i \in F_n^{av}$ و $V_{n+1}^i = V_n \cup \{v_i\}$ ، انتخاب می‌کنیم.

مرحله ۵: مجموعه داده‌های آموزش را با استفاده از $F_{n+1} = F_{n+1}^j$ و $V_{n+1} = SV_{n+1}^j$ بروز رسانی می‌کنیم.

مرحله ۶: متوسط دقت طبقه‌بندی را محاسبه می‌کنیم.

مرحله ۷: مرحله ۳ را تا جایی تکرار می‌کنیم که هیچ افزایش قابل توجهی در دقت طبقه‌بندی ایجاد نشود.

مرحله ۸: مدل ماشین بردار پشتیبان وزن دهی‌شده با حجم معاملات را با استفاده از زیر مجموعه انتخاب‌شده از روش انتخاب ویژگی هیبرید آموزش می‌دهیم تا طبقه‌بندی کننده ماشین بردار پشتیبان وزن‌دهی شده را به دست آوریم. سپس دقت طبقه‌بندی را برآورد می‌کنیم.

نمودار ۲، شمای کلی مراحل فرآیند بالا را نشان می‌دهد(لی، ۲۰۰۹).

در ادامه به صورت خلاصه، توضیحاتی در مورد شاخص‌های تکنیکالی داده می‌شود که به عنوان بردار ورودی بردار پشتیبان استفاده می‌کنیم. لازم به ذکر است که P_t^o، P_t^H، P_t^L و P_t^c به ترتیب نشان دهنده مقادیر مربوط به قیمت شروع، بالاترین قیمت، پایین‌ترین قیمت و قیمت پایان روز می‌باشند.

حجم جاری معاملات (OBV)[1]، یک شاخص آنی است که میزان افزایش و کاهش حجم معاملات را اندازه‌گیری می‌کند و روندهای حجم و قیمت را به یکدیگر مرتبط می‌سازد. این شاخص به صورت زیر محاسبه می‌شود:

اگر داشته باشیم، $P_t^C > P_{t-1}^C$، آن‌گاه داریم $OBV_t = OBV_{t-1} + V_t$

اگر داشته باشیم، $P_t^C < P_{t-1}^C$ آن‌گاه داریم، $OBV_t = OBV_{t-1} - V_t$

و اگر داشته باشیم $P_t^C = P_{t-1}^C$، آن‌گاه داریم، $OBV_t = OBV_{t-1}$.

زمانی‌که حجم معاملات افزایش می‌یابد، نه تنها قیمت جهش رو به بالا خواهد داشت و همین طور برعکس، بلکه این مساله نشان می‌دهد که سرمایه‌گذاران و حجم سرمایه زیادی در این تغییرات سهیم هستند(گرانویل، ۱۹۶۰).

تغییر قیمت یک سهم (ATR)[2] در هر روز، اطلاعاتی در مورد میزان نوسان‌های قیمت آن سهم می‌دهد. دامنه نوسان‌های واقعی، میزان نوسان‌ها را برای دوره مورد نظر محاسبه می‌کند و بر این اساس یک شاخص کمکی به نام دامنه واقعی (TR) به دست می‌آید. دامنه واقعی به صورت زیر محاسبه می‌شود:

$$TR_t = \max\{\ P_t^H - P_t^L,\ P_t^H - P_t^{c-1},\ P_t^L - P_t^{c-1}\}$$

دامنه نوسان‌های واقعی میانگین حسابی n روزه از دامنه واقعی می‌باشد.

---

1. On Balance Volume
2. Average True Range

$$ATR_t^n = \frac{1}{n\sum_{i=0}^{n} TR_{t-i}}$$

به منظور کمی کردن نوسان‌ها، دامنه متوسط واقعی بر انحراف استاندارد قیمت‌های پایانی برتری دارد. زیرا نوسان‌های قیمتی در طول روز را نیز در نظر می‌گیرد(گوستاوسون، ۲۰۰۱).

معامله‌گران از شاخص قدرت نسبی(RSI)[1]، به منظور تعیین وضعیت‌های اشباع خرید و یا فروش استفاده می‌کنند(فابر، ۱۹۹۴).

$$RSI_t^n = 100 - \frac{100}{1 - RS_t^n}$$

که در آن RS_t^n، جمع عایدی‌ها و یا ضررها در طول n روز گذشته است که در روز t ام محاسبه شده است.

$$RS_t^n = \frac{\sum_{t=0}^{n} U_t}{\sum_{t=0}^{n} D_t}$$

اگر داشته‌باشیم، $P_t^C \geq P_{t-1}^C$، آن‌گاه داریم، $U_t = P_t^C - P_{t-1}^C$

اگر داشته‌باشیم، $P_t^C < P_{t-1}^C$ آن‌گاه داریم، $D_t = P_{t-1}^C - P_t^C$.

از شاخص نوسان‌گر %R ویلیام[2] برای تعیین میزان قدرت و پایداری روند فعلی و از آن مهم‌تر برای زمان روند آماده تغییر جهت استفاده می‌شود. این شاخص تعیین می‌کند که چه زمانی دارایی در حالت اشباع خرید و یا اشباع فروش قرار دارد و به صورت زیر محاسبه می‌شود:

$$\%R = 100 * \frac{\max P_t^n - P_t}{maxP_t^n - minP_t^n}$$

که در آن داریم،

$$\text{Max}\, P_t^n = \max\{P_t, P_{t-1}, \dots, P_{t-n}\}$$

$$\text{Min}\, P_t^n = \max\{P_t, P_{t-1}, \dots, P_{t-n}\}$$

1. Relative Strength Index
2. Williams %R Oscillator

شاخص ورتکس[1] نشان دهنده تغییر جهت‌هاست. براساس ایده اصلی شاخص، اعلانات متوالی از بازار می‌تواند حاوی اطلاعاتی در مورد جهت روند باشد (بوتز و سیپمن، ۲۰۱۰). در واقع شاخص ورتکس شامل دو مقدار $V_t^{n(+)}$ و $V_t^{n(-)}$ می‌باشد که نشان دهنده تغییرات مثبت و منفی در روند قیمت می‌باشند. این مقادیر به صورت زیر محاسبه می‌شوند:

$$V_t^{n(+)} = \frac{\sum_{k=0}^{d}(P_{t-k}^{H} - P_{t-k}^{L})}{\sum_{k=0}^{d} TR_{t-k}}$$

$$V_t^{n(+)} = \frac{\sum_{k=0}^{d}(P_{t-k}^{L} - P_{t-k}^{H})}{\sum_{k=0}^{d} TR_{t-k}}$$

که در آن n تعداد روزهایی است که شاخص برای آن محاسبه شده است.

به طور ساده، عایدی یک سرمایه‌گذاری، نرخ بازده n روزه R_t^n است و ریسک مرتبط با آن، انحراف استاندارد این بازده‌ها در طول یک دوره n روزه است. این دو شاخص در طبقه‌بندی چندان سودمند و حاوی اطلاعات مهمی نیستند و با توجه به وجود فرآیند انتخاب ویژگی انتظار داریم که این دو در زیرمجموعه نهایی ویژگی‌ها برای آزمون مدل ماشین بردار پشتیبان وجود نداشته باشند.

پس از محاسبه شاخص‌ها و فراهم‌کردن بردارهای ورودی مدل، از آن‌جایی‌که دامنه تغییرات مقادیر مربوط به هر یک از این شاخص‌ها متفاوت است، به نرمال‌سازی داده می‌پردازیم. نرمال‌سازی داده‌ها دقت طبقه‌بندی کننده SVM را افزایش می‌دهد(تئودوریدیس، ۲۰۰۸). مقادیر تمامی ویژگی‌ها، x_i ، را محدود به بازه [L,U] می‌کنیم که، $L,U \in R$ و L<U می‌باشد.

$$x'_i = L + (U - L)\frac{x_i - m_i}{M_i - m_i}$$

که در آن M_i و m_i، به ترتیب نشان دهنده مقادیر بیشینه و کمینه هر ویژگی می‌باشد و به صورت زیر تعریف می‌شود:

$$M^f = \max\{x_i\}$$
$$m_i = \min\{x_i\}$$

---

1. Vortex indicator

به منظور تایید توانایی طبقه‌بندی کننده ماشین بردار پشتیبان وزن‌دهی‌شده همراه با فرآیند انتخاب ویژگی هیبرید پیشنهادی، در پیش‌بینی روند، یک استراتژی معاملاتی ارائه می‌شود. این استراتژی فعال با توجه به پیش‌بینی روند حاصل شده از طبقه‌بندی کننده، براساس شاخص‌های تکنیکال می‌باشد و خرید و فروش‌ها در زمان‌های تشخیص تغییر روند انجام می‌شوند. لازم به دکر است که پنجره زمانی معاملاتی ۱۴ روزه است.

تجزیه و تحلیل داده‌ها و آزمون فرضیه‌ها

در ادامه، نتایج پژوهش و تحلیل آن‌ها در دو بخش مربوط به نتایج فرآیند انتخاب ویژگی و طبقه‌بندی آورده شده است. فرآیند نشان داده شده در نمودار ۲، بر روی داده‌های روزانه مربوط به ۱۰ سهم انتخاب شده از سهم‌های بورس اوراق بهادار پیاده‌سازی می‌شود. از داده‌های روزانه ۸ سال نخست برای آموزش و آزمون شبکه استفاده و آزمون استراتژی معاملاتی بر روی داده‌های ۲ سال آخر انجام شده است.

مقدار K، تعداد ویژگی‌های انتخاب شده در هر زیرمجموعه جداگانه از ویژگی‌ها می‌باشد. طبق فرمول محاسبه K، مقادیر K برابر با ۲، ۴ و ۶ حالت حدی است که در آن، انتخاب ویژگی، هیچ تاثیری بر روی مجموعه ویژگی‌های اولیه ندارد. نتایج مربوط به متوسط دقت طبقه‌بندی کننده هر یک از این حالات، در جدول ۱ آمده است.

جدول ۱. مقایسه دقت نتایج برای مقادیر مختلف K

تعداد ویژگی‌ها	دقت طبقه‌بندی کننده	
	(٪)آموزش	(٪)تست
$k=\left\lfloor\frac{\lvert F\rvert}{4}\right\rfloor=۲$	٪ ۷٤/۵۰	٪ ۷۳/۵۲
$k=\left\lfloor\frac{2*\lvert F\rvert}{4}\right\rfloor=٤$	٪ ۷۳/۸۰	٪ ۷۱/۵۹
$K=\lvert F\rvert=٦$	٪ ٦٦/۲۰	٪ ٦۹/۹۱

با توجه به نتایج جدول ۱ می‌بینیم که با افزایش تعداد ویژگی‌ها، دقت طبقه‌بندی کننده کاهش می‌یابد. لازم به یادآوری است که ماشین بردار پشتیبان را به وسیله حجم معاملات واقعی وزن‌دهی کرده‌ایم.

همان‌طور که در جدول ۲ دیده می‌شود، مقدار امتیاز فیشر برای شاخص مربوط به حجم معاملات از همه بیشتر و دقت طبقه‌بندی به وسیله آن بیشتر است. علاوه بر این، همان‌طور که انتظار داشتیم، در این رتبه‌بندی، شاخص‌های نرخ بازده کلی و انحراف استاندارد، کمترین میزان امتیاز فیشر را دریافت کرده‌اند و اضافه کردن آن‌ها به مجموعه ویژگی‌ها به میزان قابل توجهی، دقت طبقه‌بندی‌کننده را کاهش می‌دهد.

طبق نتایج جدول ۱، انتخاب K= ۲، سبب بیشترین دقت می‌شود و علاوه بر این، پس از انجام فرآیند جست و جوی ترتیبی پیشرو پشتیبانی شده، استفاده از دو ویژگی حجم جاری معاملات و دامنه متوسط واقعی به بهترین نتایج می‌انجامد. نتایج امتیازات فیشر برای هر ویژگی و متوسط دقت پیش‌بینی با اضافه شدن هر ویژگی طبق فرآیند جست و جوی ترتیبی پیشروی پشتیبانی شده، به زیر مجموعه ویژگی‌ها در جدول ۲ آورده شده است.

جدول ۲. نتایج امتیازات انتخاب ویژگی فیشر و دقت متوسط طبقه‌بندی براساس مدل

ردیف	ویژگی	امتیاز فیشر	دقت متوسط (٪)
۱	OBV	۰/۰۸۲۷	٪۷۳/۵۰
۲	ATR	۰/۰۰۵٤	٪۷۳/۵٤
۳	%R	۰/۰۰۲۷	٪۷۰/۸۲
٤	RSI	۰/۰۰۰٤	٪٦۹/۵۰
۵	R	۰/۰۰۰۳	٪٦۸/۹۰
٦	σ	۰/۰۰۰۱	٪٦۹/۲۰

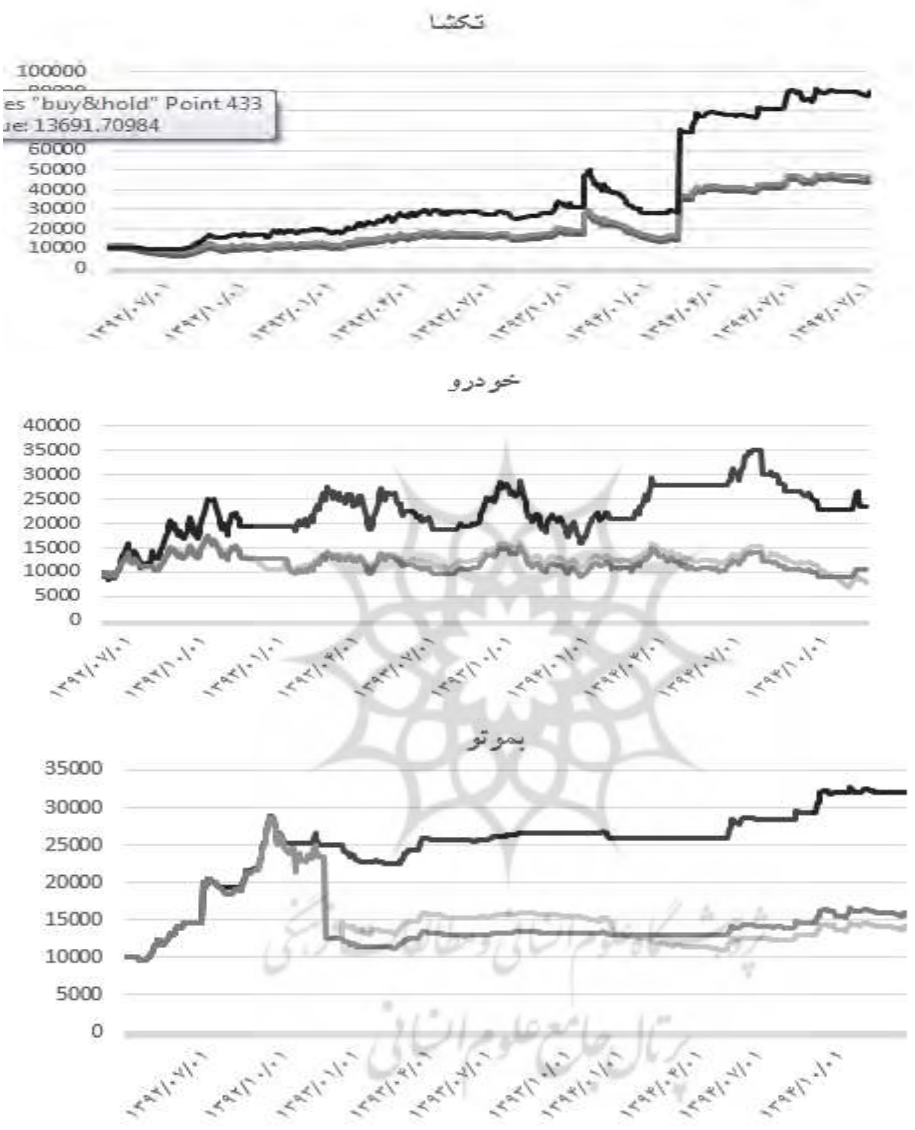
در نمودار ۳، نتایج استراتژی معاملاتی برای ۵ سهم از ۱۰ سهم انتخابی آورده شده است. نتیجه ترکیب ماشین بردار پشتیبان وزن‌دهی شده، با انتخاب ویژگی هیبرید پیشنهادی به بهبودهای قابل توجهی در نتایج استراتژی معاملاتی می‌انجامد. نتایج موجود در جدول ۳، نشان دهنده برتری قابل توجه میزان سودآوری استراتژی معاملاتی بر پایه مدل پیشنهادی نسبت به سودآوری استراتژی‌های

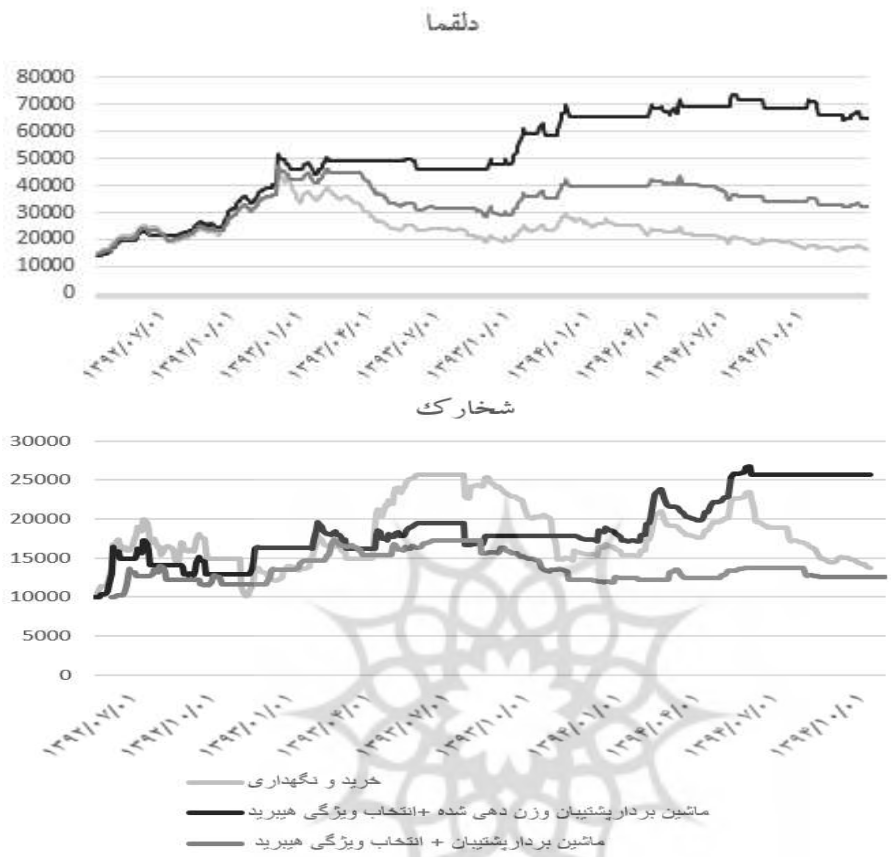
معاملاتی خرید و نگه‌داری[1] و استراتژی معاملاتی بر پایه مدل ماشین بردار پشتیبان وزن‌دهی شده با حجم معاملات اما بدون انتخاب ویژگی می‌باشد.

جدول ۳: مقایسه مقادیر بازده و بیشینه ضرر به دست آمده برای هر سهم، با استراتژی‌های مختلف

ردیف	سهم	خرید و نگهداری		استراتژی معاملاتی پیشنهادی		استراتژی ماشین بردار پشتیبان وزن‌دهی‌شده بدون انتخاب ویژگی	
		بازده ٪	بیشینه ضرر ٪(DD)	بازده ٪	بیشینه ضرر ٪(DD)	بازده ٪	بیشینه ضرر ٪(DD)
۱	تکشا	٪۴۰	٪۸۲-	٪۷۹۷	٪۶-	٪۳۴۲	٪۴۲
۲	خودرو	٪۱۲	٪۵۵-	٪۳۰۳	٪۱۸-	٪۸	٪۵۶-
۳	بموتو	٪۴۰	٪۵-	٪۲۱۹	٪۵-	٪۵۰	٪۰
۴	دلقما	٪۵۷	٪۴-	٪۵۴۷	٪۰	٪۲۱۶	٪۳-
۵	شخارک	٪۳۰۳	٪۱۰۰-	٪۴۵۸	٪۹-	٪۲۸۹	٪۱۳-
۶	ولغدر	٪۲۰۶	٪۱۹-	٪۳۶۸	٪۴-	٪۱۹۶	٪۱۸-
۷	رانفور	٪۲۸۲	٪۱۰-	٪۴۳۰	٪۰	٪۲۸۶	٪۸-
۸	سهرمز	٪۱۰۷	٪۲-	٪۲۹۸	٪۱-	٪۱۵۶	٪۳-
۹	ومعدن	٪۱۹-	٪۲۵-	٪۱۰۴	٪۱۴-	٪۶۶	٪۲۲-
۱۰	ثاژن	٪۳۲۹	٪۱۰-	٪۵۴۳	٪۱-	٪۱۹۶	٪۱۰-
میانگین		٪۱۳۶	٪۳۱-	٪۴۰۷	٪۶-	٪۱۸۱	٪۹-

1. Buy and hold


تکشا
es "buy&hold" Point 433
ue: 13691.70984
100000
60000
50000
40000
30000
20000
10000
0
خودرو
40000
35000
30000
25000
20000
15000
10000
5000
0
پمو تو
35000
30000
25000
20000
15000
10000
5000
0



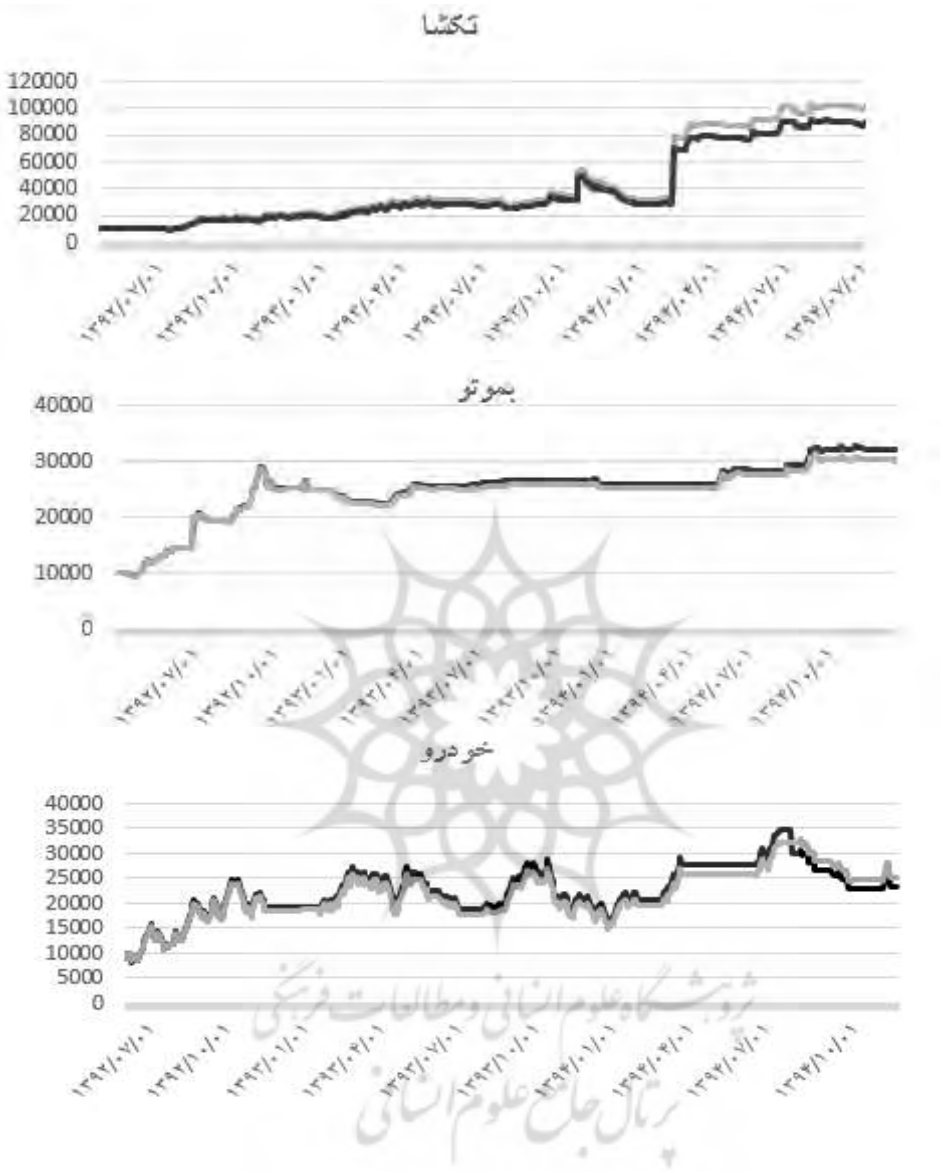
نمودار۳: مقایسه نتایج استراتژی معاملاتی با طبقه‌بندی کننده‌های مختلف و نیز استراتژی خرید و نگه‌داری

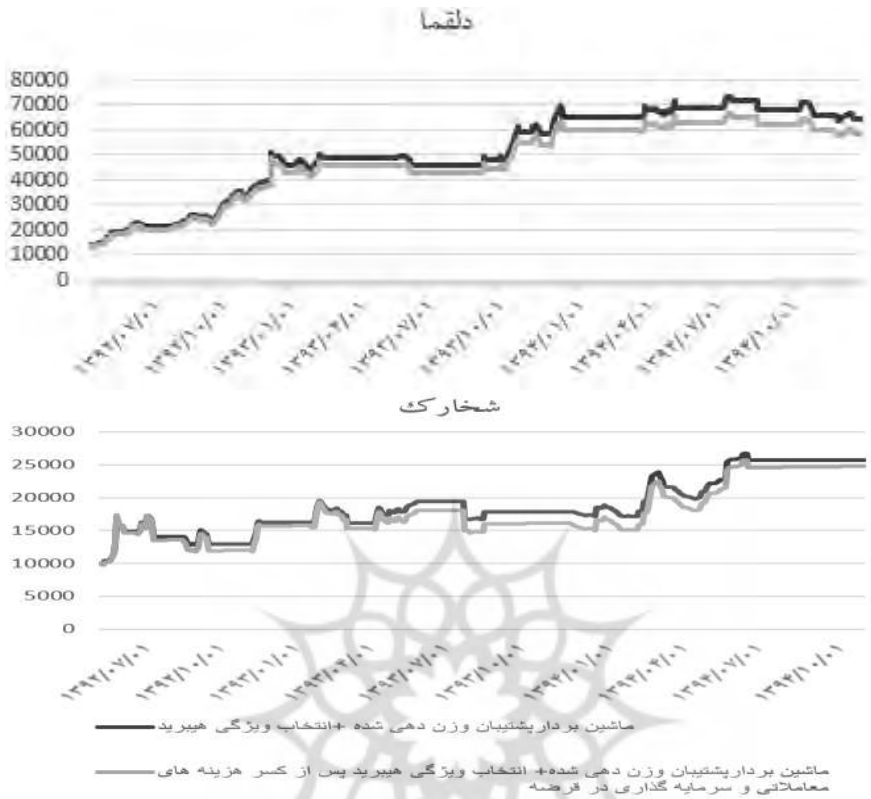
توجه داشته باشید که معاملات سهام در بازارهای واقعی همراه با هزینه‌های معاملاتی برای هر دو طرف خریدار و فروشنده می‌باشد. این هزینه‌ها شامل کارمزدهای کارگذاری، کارمزدهای بورس، کارمزد حق نظارت سازمان، کارمزد شرکت سپرده‌گذاری، کارمزد شرکت فناوری و مالیات نقل و انتقال می‌باشد. به طور کلی در هر معامله، خریدار به اندازه ۰/۴۸۶ درصد ارزش معامله و فروشنده به اندازه ۱/۰۲۹ درصد ارزش معامله، کارمزد می‌پردازد. این مبالغ به ویژه در معاملات پربسامد بر سودآوری سرمایه‌گذاری، اثرهای چشم‌گیری دارند. بنابراین لازم است به منظور نزدیک‌کردن نتایج پژوهش به واقعیت، این هزینه‌ها در استراتژی معاملاتی بهینه پیشنهادی لحاظ شوند. همچنین با توجه با این‌که برپایه نتایج حاصل از مدل پیش‌بینی، در دوره‌های نسبتا طولانی

خارج از بازار هستیم و سرمایه درگیر سهام نیست، می‌توانیم به منظور سودآوری بیشتر، دارایی خود را طی چنین دوره‌هایی در اوراق قرضه سرمایه‌گذاری کنیم. در این‌جا سود اوراق قرضه ۲۰٪ به صورت روز شمار در نظرگرفته شده است. در نمودار ۴، نتایج استراتژی معاملاتی بر پایه انتخاب ویژگی هیبرید و ماشین بردار پشتیبان وزن‌دهی شده، با و بدون در نظر گرفتن هزینه‌های معاملاتی و سود سرمایه‌گذاری در اوراق قرضه، آورده‌شده است. در جدول ۴، مقادیر کمی مربوط به تاثیر در نظرگرفتن این هزینه‌ها در سودآوری استراتژی معاملاتی آورده شده است.

جدول ٤: مقایسه مقادیر بازده و بیشینه ضرر به دست آمده برای هر سهم، با استراتژی پیشنهادی، با و بدون درنظرگرفتن هزینه‌های معاملاتی

ردیف	سهم	استراتژی معاملاتی پیشنهادی بدون هزینه‌های معاملاتی و سرمایه‌گذاری در اوراق قرضه		استراتژی معاملاتی پیشنهادی با در نظر گرفتن هزینه‌های معاملاتی و سرمایه‌گذاری در اوراق قرضه	
		بازده ٪	بیشینه ضرر ٪(DD)	بازده ٪	بیشینه ضرر ٪(DD)
۱	تکشا	٪۷۹۷	٪٦-	٪۹۰۹	٪۰
۲	خودرو	٪۳۰۳	٪۱۸-	٪۳۰۷	٪۱۸-
۳	بموتو	٪۲۱۹	٪۵-	٪۱۹۹	٪۵-
٤	دلقما	٪٥٤٧	٪۰	٪٤٨٣	٪۱-
٥	شخارک	٪٤٥٨	٪۹-	٪٥۲۱	٪۹-
٦	ولغدر	٪۳٦٨	٪٤-	٪۳۷۱	٪٤-
۷	رانفور	٪٤۳۰	٪۰	٪۳٤٥	٪۰
۸	سهرمز	٪۲۹۸	٪۱-	٪۲۲۱	٪۱-
۹	ومعدن	٪۱۰٤	٪۱٤-	٪۱۸٥	٪۱۰-
۱۰	ثاژن	٪٥٤۳	٪۱-	٪٤۳۲	٪۱-
میانگین		٪٤۰۷	٪٦-	٪۳۹۷	٪٥-


تکشا
120000
100000
80000
60000
40000
20000
0
بموتو
40000
30000
20000
10000
0
خودرو
40000
35000
30000
25000
20000
15000
10000
5000
0



نمودار ٤- مقایسه نتایج استراتژی بهینه با و بدون کسر هزینه‌های معاملاتی و سرمایه‌گذاری در اوراق قرضه

نتیجه‌گیری و بحث

در این پژوهش مدل ترکیبی جدید از ماشین بردار پشتیبان وزن‌دهی شده با حجم معاملات واقعی و انتخاب ویژگی هیبرید، آورده شد. همچنین به منظور اثبات توانایی و قابلیت مدل پیشنهادی در پیش‌بینی درست و قابل اطمینان نوسان‌ها، یک استراتژی معاملاتی برپایه مدل طبقبندی کننده و شاخص‌های تکنیکال داده شد که توانست برتری دقت و عملکرد مدل پیشنهادی را بر مدل ماشین بردار پشتیبان وزن‌دهی‌شده بدون انتخاب ویژگی و همچنین استراتژی خرید و نگه‌داری ثابت‌کند. همچنین لازم به یادآوری است که در این پژوهش از معتبرسازی مقطعی استفاده شده است و در هربار، اجرای الگوریتم از ۸۰ درصد داده‌های ۸سال نخست، برای آموزش و ۲۰ درصد برای آزمون استفاده شده و استراتژی معاملاتی بر روی داده‌های ۲ سال آخر پیاده‌سازی شده است. بنابراین،

می‌توان ادعا کرد که دقت‌های پیش‌بینی روش‌های مختلف در جداول داده شده و همچنین نتایج استراتژی، قابل اطمینان می‌باشد.

در مجموع، دقت طبقه‌بندی با روش ماشین بردار پشتیبان وزن‌دهی شده با حجم معاملات واقعی و انتخاب ویژگی هیبرید ۷۳/۵ درصد و با روش ماشین بردار پشتیبان وزن‌دهی‌شده، بدون انتخاب ویژگی ۶۹/۹ درصد می‌باشد. این مقادیر در پژوهش (مینگ‌لی، ۲۰۰۹) که از مدلی با ویژگی‌ای مشابه با مدل ما همراه با ماشین بردار پشتیبان ساده دارد، به ترتیب ۸۵/۵ درصد و ۸۳/۲ درصد اعلام شده است. علت بالاتر بودن دقت این پژوهش نسبت به پژوهش ما را می‌توان در دقت کمتر و در مواردی فقدان داده‌های معاملاتی سهام بورس اوراق بهادار تهران دانست. همچنین نتایج نشان می‌دهد که شاخص‌های حجم جاری معاملات و دامنه متوسط واقعی با دقت قابل قبولی، توانایی پیش‌بینی قیمت سهم را دارند. این مورد با وزن‌دهی شده بودن ماشین بردار پشتیبان با حجم معاملات هم‌خوانی دارد.

متوسط بازده سالانه کسب شده توسط استراتژی معاملاتی پیشنهادی در پژوهش (زیبکوسکی، ۲۰۱۴) که در آن از مدل ماشین بردار پشتیبانی، مشابه مدل مورد استفاده این پژوهش، استفاده شده است، ۱۱۸/۷۱ درصد می‌باشد. در این پژوهش با توجه به نتایج داده شده، بازدهی ده سهم مورد بررسی در طول دو سال ۹۳ و ۹۴، برای استراتژی خرید و نگهداری به طور متوسط برابر با ۵۳٪، برای استراتژی داده شده با استفاده از ماشین بردارپشتیان بدون انتخاب ویژگی برابر با ۶۷٪ می‌باشد. در حالی که با استفاده از استراتژی معاملاتی پیشنهادی به طور متوسط در هر سال سود ۱۲۳/۶۰ درصدی خواهیم داشت. همچنین پس از در نظرگرفتن هزینه‌های معاملاتی و خرید اوراق قرضه در روزهایی که موقعیت هر سهم بسته می‌باشد، سود متوسط سالانه برابر با ۱۲۲/۹ درصد می‌باشد که این مقایسه نشان می‌دهد، میزان بازدهی کسب شده با استفاده از روش پیشنهادی در بورس اوراق بهادار تهران، قابل مقایسه با دقت به کارگیری این روش در سطح بین‌المللی است.

با توجه به این پژوهش می‌توان پیشنهادهایی برای پژوهش‌های آتی به شرح زیر داد:

از جمله مواردی که هنوز جای مطالعه و بررسی دارد، تابع وزن‌دهی به تابع جریمه در مدل طبقه‌بندی‌کننده ماشین بردار پشتیبان می‌باشد. در واقع در این‌جا تابع جریمه به حجم معاملات روزانه حساس شده است. می‌توان این تابع جریمه را به نحوی تغییر داد که به نوسان‌ها حساس باشد. یک تابع جریمه پیشنهادی می‌تواند به صورت $\frac{TR_t}{\sum_{i=0}^{n} TR_{t-i}}$ باشد.

همچنین بررسی سایر روش‌های انتخاب ویژگی به منظور افزایش دقت پیش‌بینی، از جمله روش جنگل تصادفی[1] به منظور مشخص کردن اهمیت ویژگی‌ها، می‌تواند محل مطالعات بیشتر واقع‌شود (نمونه‌ای از این کار در پژوهش چن و لین ۲۰۰۵، آمده است). علاوه بر این می‌توان در بخش فیلتر کننده الگوریتم هیبرید پیشنهادی، تغییراتی ایجاد کرد و از روش‌هایی مانند بهره اطلاعات[2]، عدم قطعیت متقارن[3] و انتخاب ویژگی برپایه همبستگی[4] استفاده نمود.

همچنین گسترش مجموعه داده‌های آموزش می‌تواند حوزه جالبی برای بهبود مدل و عملکرد آن باشد.

علاوه بر این موارد، به منظور بررسی عملکرد الگوریتم پیشنهادی می‌توان از داده‌های غیر بورسی و یا بازارهای بورس خارج از کشور، به ویژه بازارهای نوظهور استفاده نمود و این الگوریتم را در این بازارها مورد آزمون و مطالعه قرار داد.

- همچنین می‌توان از تابع جریمه با استفاده از حجم روزانه معاملات، در انواع دیگری از الگوریتم‌های یادگیری ماشینی، از جمله شبکه‌های عصبی و یا مجموعه‌های فازی استفاده‌کرد و نتایج حاصل را با نتایج به دست‌آمده در این پژوهش مقایسه نمود.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Random forest
2. Information gain
3. Symmetrical uncertainty
4. Correlation based feature selection

منابع

- سینایی حسنعلی، مرتضوی سعید الله، تیموری اصل یاسر."پیش‌بینی شاخصبورس اوراق بهادار تهران با استفاده از شبکه‌های عصبی"، *بررسی‌های حسابداری و حسابرسی*، شماره ۴۱، دوره ۱۲، صص. ۵۹ - ۸۳.
- عبادی،ا."پیش‌بینی قیمت شاخص کل سهام در بازار بورس تهران با استفاده از شبکه‌های عصبی مصنوعی". *پایان‌نامه کارشناسی ارشد*، دانشکده اقتصاد و علوم اجتماعی دانشگاه بوعلی سینا، همدان.
- فلاحپور، سعید، گل ارضی، حسین، فتوره چیان، ناصر."پیش‌بینی روند حرکتی قیمت سهام با استفاده از ماشین بردار پشتیبان بر پایه ژنتیک در بورس اوراق بهادار تهران"، *تحقیقات مالی*، شماره ۲، دوره ۱۵، صص. ۲۶۹-۲۸۸.
- فلاحپور، سعید، طبسی، ملیحه، "برآورد ارزش در معرض ریسک با استفاده از مدل ترکیبی ماشین بردار پشتیبان و گارچ"، راهبرد مدیریت مالی الزهرا، شماره ۱، دوره ۱، صص ۱۷۷-۱۹۵.
- منجمی، سید امیر حسین، ابزاری، مهدی، رعیتی شوازی، علیرضا."پیش‌بینی قیمت سهام در بازار بورس اوراق بهادار با استفاده از شبکه عصبی مصنوعی". *فصلنامه اقتصاد مالی*، شماره ۳، دوره ۶، صص.۱-۲۶.
- هاشمی احمد. "تاثیر فاکتورهای رفتاری بر پیش‌بینی قیمت سهام با استفاده از مدل شبکه‌های عصبی رگرسیونی جلوسو"، پایان نامه کارشناسی ارشد. دانشکده صنایع دانشگاه علم و فرهنگ، تهران.
- Abraham, A., Nath, B., & Mahanti, P. K. (2001) "Hybrid intelligent systems for stock market analysis." In , pp.337-345.
- Bao, D. Yang, Z. (2008). Intelligent stock trading system by turning point confirming and probabilistic reasoning, , 34(1), pp. 620–627.
- Blum AL, Langley P. (1997). Selection of relevant features and examples in machine learning. 97(1), pp.245–70.
- Botes, E., & Siepman, D. (2010). The Vortex Indicator. , 28(1), p.21.
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. 40(1), pp. 16-28.

- Chavarnakul, T., & Enke, D. (2008) Intelligent technical analysis based equivolume charting for stock trading using neural networks. pp.1004-1017.
- Chen, Y.-W., & Lin, C.-J. (2006). Combining SVMs with various feature selection strategies. , Springer Berlin Heidelberg, pp. 315-324.
- Chong, E. K., & Zak, S. H. (2013). . Vol.76. John Wiley & Sons.
- Choudry, R. & Grag, K. (2008). A Hybrid Machine Learning System for Stock Market Forecasting. (3), pp.315-318.
- Fodor, I. K. (2002). A survey of dimension reduction techniques.
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. ; 46(1-3), pp.389–422.
- Granville, J. E. (1960). A strategy of daily stock market timing for maximum profit". .
- Gustafson, G. (2001). Which Volatility Measure? Is average true range, an approximation, superior to standard deviation, the most beloved of quants, as a measure of volatility?. 19(6), pp. 46-50.
- Huang, S. C., & Wu, T. K. (2008). Integrating GA-based time-scale feature extractions with SVMs for stock index forecasting. s, 35(4), pp. 2080-2088.
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. 97(1), pp.273-324.
- Langley P. (1994, November). Selection of relevant features in machine learning. In , Vol. 184, pp. 245-271
- Lee, M. C. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. pp. 10896-10904.
- Leung, M. T., Chen, A. S., & Daouk, H. (2000). Forecasting exchange rates using general regression neural networks. pp.1093-1110
- Liu, H., & Setiono, R. (1996, July). A probabilistic approach to feature selection-a filter solution. In , Vol. 96, pp. 319-327.
- Liu, Y., & Zheng, Y. F. (2006). (2006). FS_SFS: A novel feature selection method for support vector machines. pp.1333-1345.
- Malkiel, B. G., & Fama, E. F. (1970), Efficient capital markets: A review of theory and empirical work, , 25(2), pp.383–417.

- Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters, 28(4), pp. 603-614.
- Nair, B.B., Mohandas, V.P. & Sakthivel, N.R. (2010). A Genetic Algorithm Optimized Decision Tree-SVM based Stock Market Trend Prediction System. . 2 (9), pp. 2981-2988.
- Narendra P, Fukunaga K. (1977). A branch and bound algorithm for feature subset selection. , 100(9), pp.917–22.
- Pudil, P., Novovičová, J., & Kittler, J. (1994), Floating search methods in feature selection. ; 15(11), pp.1119–25.
- Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. (19), pp.2507-2517
- Teixeira, L. A., & De Oliveira, A. L. I. (2010). A method for automatic stock trading combining technical analysis and nearest neighbor classification. , 37(10), pp.6885-6890.
- Theodoridis, S., & Koutroumbas, K. (1999). Pattern Recognition.
- Wang, L. (Ed.). (2005). Support vector machines: theory and applications. ,Vol. 177
- Żbikowski, K. (2015). Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. (4), pp.1797-1805
- Zhang, X., Hu, Y., Xie, K., Wang, S., Ngai, E. W. T., & Liu, M. (2014). A causal feature selection algorithm for stock prediction modeling. 142 pp. 48-59.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی