تحقیقات مالی
دانشکدهٔ مدیریت دانشگاه تهران

دورهٔ ۱۵، شمارهٔ ۲
پاییز و زمستان ۱۳۹۲
ص. ۲۸۸- ۲۶۹

پیش‌بینی روند حرکتی قیمت سهام با استفاده از ماشین بردار پشتیبان برپایهٔ الگوریتم ژنتیک در بورس اوراق بهادار تهران

سعید فلاح‌پور[۱]، غلامحسین گل‌ارضی[۲]، ناصر فتوره‌چیان[۳]

چکیده: با توجه به گسترش روزافزون روش‌های پیش‌بینی در بازارهای مالی و نیز، از آنجا که قیمت سهام یکی از مهم‌ترین عوامل مؤثر در تصمیمات سرمایه‌گذاری است و پیش‌بینی آن می‌تواند نقش با اهمیتی در این زمینه ایفا کند، در این پژوهش سعی شده است، مدلی ارائه شود تا بر اساس آن بتوان روند حرکتی قیمت سهام شرکت مورد نظر را با دقت بالایی پیش‌بینی کرد. بر همین اساس، یک مدل ترکیبی برای پیش‌بینی روند حرکتی قیمت سهام با استفاده از ماشین بردار پشتیبان بر پایهٔ الگوریتم ژنتیک ارائه شده است. برای نمونهٔ آماری، سی شرکت از پنجاه شرکت برتر بورس اوراق بهادار در سه ماههٔ دوم سال ۹۰ انتخاب شده است. سپس برای هر سی شرکت، ۴۴ متغیر محاسبه شد. این متغیرها ورودی مدل ترکیبی هستند و به‌کمک الگوریتم ژنتیک بهینه‌سازی شده‌اند. نتایج نشان می‌دهد، مدل ترکیبی ماشین بردار پشتیبان بر پایهٔ الگوریتم ژنتیک در پیش‌بینی روند حرکتی قیمت سهام بسیار بهتر عمل کرده و درمقایسه با روش ماشین بردار پشتیبان ساده، از دقت بالاتری برخوردار است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

واژه‌های کلیدی: پیش‌بینی، تحلیل تکنیکی، قیمت سهام، الگوریتم ژنتیک، ماشین بردار پشتیبان.

۱. استادیار، مدیریت مالی، دانشگاه تهران، ایران
۲. استادیار، مدیریت مالی، دانشگاه سمنان، سمنان، ایران
۳. کارشناس ارشد، MBA گرایش مدیریت مالی، دانشگاه سمنان، سمنان، ایران

تاریخ دریافت مقاله: ۱۳۹۲/۰٤/۳۰
تاریخ پذیرش نهایی مقاله: ۱۳۹۲/۱۰/۱۷
نویسندهٔ مسئول مقاله: ناصر فتوره‌چیان
E-mail: naserfatourechian@gmail.com

مقدمه

پیش‌بینی و بررسی رفتار قیمت اوراق بهادار، مقوله‌ای است که دانشمندان علوم مالی و سرمایه‌گذاران همواره در پی آن هستند. دلیل اصلی سرمایه‌گذاری در بازار سهام، به‌دست آوردن سود است که لازمهٔ آن، داشتن اطلاعات درست از بازار بورس و تغییرات سهام و پیش‌بینی روند آینده آن است؛ بنابراین، سرمایه‌گذار نیازمند ابزارهای قدرتمند و قابل‌اعتماد است تا از طریق آن به پیش‌بینی قیمت سهام بپردازد (شینا و تانگ، ۲۰۰۸).

با توجه به افزایش تعداد سهامداران در بازار سهام و توجه آنها به قیمت سهام مختلف در معاملات، پیش‌بینی روند حرکتی قیمت سهام اهمیت بسزایی یافته است. بسیاری از افراد در زمان سرمایه‌گذاری، هنگام مقایسهٔ سهام مختلف، از روند حرکتی قیمت سهم استفاده می‌کنند و همچنین خواهان پیش‌بینی این روند هستند تا بدانند روند افزایشی یا کاهشی قیمت سهم مورد نظر حدوداً تا چه زمانی ادامه می‌یابد (عبده تبریزی و جوهری، ۱۳۷۵).

پژوهش‌ها نشان می‌دهد، استفاده از ابزارها و روش‌های پیش‌بینی سنتی، خطای بالایی دارد و بیشتر در مقایسه با روش‌های جدیدتر و مدل‌های غیرخطی عملکرد ضعیف‌تری دارند. در این پژوهش نیز، یکی از روش‌های هوش مصنوعی به نام ماشین بردار پشتیبان، به‌همراه یکی از الگوریتم‌های پرکاربرد در این حوزه، برای پیش‌بینی روند حرکتی قیمت سهام به‌کار گرفته شده است (ییم، ۲۰۰۲).

با پیشرفت‌های اخیر در هوش مصنوعی، روش‌های جدیدی برای پیش‌بینی ارائه شده است که نسبت به روش‌های سنتی، از دقت بالاتری برخوردار هستند. رایج‌ترین این روش‌ها، الگوریتم شبکه‌های عصبی مصنوعی است. با این حال، شبکه‌های عصبی کاستی‌هایی دارند، مانند نیاز به پارامترهای کنترلی زیاد، دشواری رسیدن به نتیجه‌ای پایدار و غیره. به‌دلیل وجود چنین ضعف‌هایی، روش‌های بهتری برای بهبود مدل شبکه‌های عصبی[1] طراحی شده است (شنگ، هسیه، چای و هسو، ۲۰۰۹). این روش‌ها با استفاده از قابلیت یادگیری خود، هرگونه تغییر ایجادشده در قوانین نهفته در سری‌های زمانی را فراگرفته، برای پیش‌بینی آینده از آن استفاده می‌کنند. ماشین بردار پشتیبان (SVM)[2] یکی از روش‌های یادگیری با نظارت است که در طبقه‌بندی و رگرسیون به‌کار برده می‌شود. این روش از جمله روش‌های به نسبت جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر طبقه‌بندی، از جمله شبکه‌های عصبی نشان داده است (وپنیک، ۱۹۹۵). الگوریتم ژنتیک[3] یک روش بهینه‌سازی الهام‌گرفته از

1. Neural Network
2. Support Vector Machines (SVM)
3. Genetic Algorithm (GA)

طبیعت جاندار است که می‌توان آن را یک روش جست‌وجوی عددی، مستقیم و تصادفی معرفی کرد (گلدبرگ، ۱۹۹۸). در پژوهش پیش رو با استفاده از مدل‌های یادشده و به‌کمک متغیرهای ورودی که ۴۴ مورد هستند، به بررسی عملکرد دقت پیش‌بینی مدل‌های یادشده پرداخته شده است. هدف اصلی این پژوهش، بررسی کارایی مدل ترکیبی در مقابل مدل ماشین بردار پشتیبان ساده است. همچنین با استفاده از الگوریتم ژنتیک در مدل ترکیبی، اثر ترکیب بهینه متغیرهای ورودی بر نتیجهٔ کلی بررسی می‌شود.

پیشینهٔ نظری پژوهش

ماشین بردار پشتیبان

فیشر اولین الگوریتم را برای طبقه‌بندی و دسته‌بندی الگوها در سال ۱۹۳۶ ارائه کرد و از طریق کم‌کردن خطای طبقه‌بندی داده‌های آموزشی، برای بهینه‌کردن الگوها به‌کار گرفت. در سال ۱۹۶۵، پژوهشگری روسی به نام ولادیمیر وپنیک، گامی بسیار مهم در طراحی طبقه‌بندی‌کننده‌ها برداشت. وی نظریهٔ آماری یادگیری را به‌صورت مستحکم‌تری بنا نهاد و ماشین‌های بردار پشتیبان را بر این اساس ارائه داد (یاری و خانلو، ۱۳۸۷).

در واقع ریشهٔ SVMها در تئوری یادگیری آماری[1] است و کاربردهای فراوانی در رگرسیون، طبقه‌بندی، خوشه‌بندی[2] و به‌طورکلی تقریب توابع دارند. رویکرد SVM در بدو ابداع، تنها طبقه‌بندی دوکلاسه را شامل می‌شد. این رویکرد در ادامه با استفاده از انواع تکنیک‌های ترکیب به طبقه‌بندی چندکلاسه نیز تعمیم یافت (آبه، ۲۰۰۵).

SVMها داده‌ها را به‌صورت بردار پردازش می‌کنند و همواره در میان تمام ابرصفحاتی که داده‌ها را تفکیک می‌کنند، آن اَبَرصفحه را برمی‌گزینند که بیشترین تفکیک‌پذیری یا بیشترین اندازهٔ حاشیه را میان داده‌های کلاس‌های مختلف حاصل کنند. ابرصفحه مطلوب طوری انتخاب می‌شود که فاصلهٔ آن از نزدیک‌ترین داده بیشینه شود. به نزدیک‌ترین نقاط آموزشی به حداکثر حاشیهٔ ابرصفحه، بردارهای پشتیبان اطلاق می‌شود. چنین ابرصفحه‌ای در صورت وجود، ابرصفحهٔ با حاشیهٔ بیشینه خوانده می‌شود (گان، ۱۹۹۸).

اگر نقاط آموزشی را به‌صورت $[x_i , y_i]$، بردار ورودی را $x_i \in Rn$ و ارزش طبقه را $y_i \in \{-1,1\}$, $i = 1,\dots,l$ تعریف کنیم، آنگاه در حالتی که داده‌ها به‌صورت خطی قابل

1. Statistical Learning Theory
2. Clustering

تفکیک هستند، قواعد تصمیم‌گیری تعریف می‌شود و با یک صفحهٔ بهینه که طبقات تصمیم‌گیری باینری را تفکیک می‌کند، به‌صورت رابطهٔ ۱ خواهد بود.

رابطهٔ ۱) $$y = sign\left(\sum_{i=1}^{N} y_i\, \alpha_i (X.X) + b\right)$$

که در آن؛ y: خروجی معادله، y_i: ارزش طبقهٔ نمونهٔ آموزشی X_i و . نشان‌دهندهٔ ضریب داخلی است. بردار $x = (x_1, x_2, \dots, x_n)$ نشان‌دهندهٔ یک دادهٔ ورودی و بردارهای Xi $X_i : i = 1, \dots, N$، بردارهای پشتیبان هستند. در رابطهٔ ۱ پارامترهای α_i , b، تعیین‌کنندهٔ ابرصفحه هستند.

اگر داده‌ها به‌صورت خطی قابل تفکیک نباشند، رابطهٔ ۱ به‌صورت رابطهٔ ۲ تغییر می‌یابد.

رابطهٔ ۲) $$Y = sign\left(\sum_{i=1}^{N} y_i\, \alpha_i K(X, X_i) + b\right)$$

تابع $K(X, X_i)$ تابع کرنلی است که برای ایجاد ماشین‌هایی با انواع مختلفی از سطوح تصمیم‌گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند (راعی و فلاح‌پور، ۱۳۸۷). برای مثال سه نوع تابع کرنل که در مدل SVM به‌کار می‌روند، عبارتند از: تابع کرنل چندجمله‌ای، تابع کرنل پایهٔ شعاعی و تابع کرنل پرسپترون چندلایه. قابلیت به‌کارگیری کرنل‌ها[1] یکی از نقاط قوت کلیدی SVMها محسوب شده، سبب می‌شود که آنها انعطاف بالایی برای حل انواع مختلف مسائل با دشواری‌های گوناگون داشته باشند (کریستیانی و تیلور، ۲۰۰۰).

برخی ویژگی‌های ماشین‌های بردار پشتیبان عبارتند از: آموزش ماشین بردار پشتیبان به آسانی صورت می‌گیرد؛ با نمونه‌های آموزشی کم، نتایج خوبی را ایجاد می‌کند؛ برخلاف شبکه‌های عصبی نیاز به بهینه‌یابی محلی ندارد و بهترین مدل را پیدا می‌کند. در مقابل این مزیت‌ها نیز باید در انتخاب تابع هسته‌ای مناسب دقت کرد (آبه، ۲۰۰۵).

الگوریتم ژنتیک

افزایش روافزون سرعت محاسبات به‌کمک رایانه، روش طراحی بهینهٔ پارامترها در مسائل را از طریق شبیه‌سازی روند تکاملی امکان‌پذیر کرده است. امروزه، روش‌های برنامه‌نویسی تکاملی و ازجمله آنها الگوریتم ژنتیک، به روش‌های بهینه‌سازی در مسائل پیش‌بینی جایگاه خاصی داده است (چونگ و زاک، ۲۰۰۱). الگوریتم‌های ژنتیک که دستهٔ خاصی از الگوریتم‌های تکاملی هستند، به دفعات در حل مسائل دشوار علوم و مهندسی استفاده می‌شوند. می‌توان گفت، این

1. Kernels

دسته الگوریتم‌ها روش مناسبی برای رسیدن به حل مسائل دارای پیچیدگی محاسباتی، در یک زمان مناسب هستند. در هر زمان می‌توان الگوریتم را متوقف‌کرده، جواب به‌دست آمده تا آن تکرار الگوریتم را استخراج کرد. چون الگوریتم دارای خاصیت تکاملی است، چنانچه خطای موجود در جواب به‌دست آمده پذیرفته نشود، می‌توان الگوریتم را ادامه داد تا به جوابی قانع‌کننده برسد. زمان رسیدن به این جواب چندان قابل پیش‌بینی نیست، اما با اجرای مکرر الگوریتم و استفاده از آمار و علم مربوط به آن، می‌توان تقریب زمان رسیدن این الگوریتم تکاملی به جواب قانع‌کننده را حدس زد. الگوریتم‌های ژنتیکی، نیاز به یافتن یک الگوریتم خاص برای مسئله را منتفی می‌کنند و تنها کار لازم برای حل مسئله، مدل‌سازی و مقداردهی صحیح به پارامترهای الگوریتم ژنتیک است. الگوریتم‌های ژنتیکی همان‌گونه که از نام آنها پیداست، دید خود را از ژنتیک و مسائل مربوطه در این حوزه گرفته‌اند. الگوریتم با ساخت یک‌سری رشته به نام کروموزوم و با انجام عملیات مشخص ژنتیکی، انتخاب، ترکیب و جهش در طول حیات رشته‌ها، به سمت تکامل و بهتر شدن آنها گام برمی‌دارد (گلدبرگ، ۱۹۸۹).

پیشینهٔ تجربی

روند مطالعات نشان می‌دهد، هرچند مدل‌های آماری توانسته‌اند پیش‌بینی‌های خوبی را درمورد قیمت سهام ارائه کنند، در عین حال مفروضات محدودکنندهٔ برخی از این مدل‌ها بر اثربخشی این روش‌ها مؤثر بوده است؛ بنابراین، به‌تدریج روش‌های دیگری برای مقابله با این محدودیت‌ها و بهبود عملکرد پیش‌بینی‌ها معرفی شد. در پژوهش‌های داخلی و خارجی مطالعات زیادی انجام شده که برتری روش ماشین بردار پشتیبان را به روش‌های پیشین، از جمله روش شبکه‌های عصبی و نیز روش‌های سنتی نشان می‌دهد. از این رو، تنها به ذکر چند نمونه از مطالعاتی پرداخته می‌شود که اخیراً در داخل و خارج از کشور انجام گرفته و بیشترین ارتباط را با موضوع پژوهش دارند.

در پژوهشی با استفاده از ترکیب ماشین بردار پشتیبان و الگوریتم ژنتیک، به پیش‌بینی روند حرکتی قیمت سهام پرداخته شده است. نتایج نشان داد، مدل ترکیبی در مقایسه با مدل سادهٔ ماشین بردار پشتیبان از دقت بالاتری برخوردار است (چودری و گری، ۲۰۰۸).

در مطالعه دیگری برای پیش‌بینی شاخص سهام، از یکپارچه‌سازی الگوریتم ژنتیک برپایهٔ ویژگی‌هایی در مقیاس زمان[1] با ماشین بردار پشتیبان استفاده کردند که یک روش رگرسیونی

1. Integrating GA-based Time-Scale Feature

ناپارامتریک به‌شمار می‌آید و در مقایسه با روش‌های شبکه‌های عصبی، SVM ساده یا مدل‌های سنتی GARCH نتیجهٔ بهتری داشت (شیان و تانگ، ۲۰۰۸).

پژوهشگران با ترکیب رگرسیون ماشین بردار پشتیبان[1] و SOM [2] که یک الگوریتم دسته‌بندی است، به پیش‌بینی قیمت سهام پرداخته‌اند. با توجه به نتایج به‌دست آمده مدل دومرحله‌ای یادشده از مدل ساده SVR عملکرد بهتری داشته است (شنگ، هسیه، چای و هسو،۲۰۰۹).

در ادامه پژوهش‌ها، از روش ترکیبی رگرسیون ماشین بردار پشتیبان و SOFM[3] استفاده کرده‌اند که به‌کمک یک روش فیلترشده برای بهتر شدن دقت پیش‌بینی انجام می‌شود. ابتدا به کمک فیلتر کردن، ویژگی‌های مهم ورودی را انتخاب‌کرده، سپس با SOFM نمونه‌های آموزشی دسته‌بندی می‌شود و SVR شاخص قیمتی سهام را پیش‌بینی می‌کند. نتایج نشان داده است که روش ترکیبی از SVR ساده، بهتر است (چنگ و تی‌سای، ۲۰۰۹).

پژوهشگر دیگری برای پیش‌بینی روند حرکتی بازار سهام از SVM و F-SSFS[4] استفاده کرده است. نتایج نشان داد این روش از روش‌های دیگر، از جمله شبکه‌های عصبی، عملکرد بالاتری دارد (مینگ، ۲۰۰۹).

چند پژوهشگر برای پیش‌بینی روند حرکتی بازار سهام، از الگوریتم ژنتیک برپایهٔ درخت تصمیم[5] استفاده کردند که از SVM برای پیش‌بینی بهره می‌برد. نتایج حاکی از برتری مدل ترکیبی در مقایسه با شبکه‌های عصبی بوده است. در این پژوهش از متغیرهای تحلیل تکنیکال به‌منزلهٔ ورودی مدل استفاده شده است (بینوی، مهندس و ساکتهویل،۲۰۱۰).

در پژوهشی در ترکیه، روش SVM و شبکه‌های عصبی برای پیش‌بینی روند حرکتی شاخص قیمتی سهام در بازار سهام استانبول به‌کار رفته است. این روش به کمک ده نماگر تکنیکال انجام شده است. نتایج حاکی از بالا بودن میانگین دقت پیش‌بینی با روش شبکه‌های عصبی نسبت به روش SVM بوده است (یاکوب، ملک و عمر، ۲۰۱۱).

پژوهشگرانی برای پیش‌بینی یک روز آتی شاخص قیمتی سهام، از رویکرد SVM با ترکیب روش انتخاب ویژگی‌های فرکتال[6] که برای حل مشکلات غیرخطی مناسب است و در انتخاب

---

1. Support Vector Regression (SVR)
2. Self-Organizing Map (SOM)
3. Self-Organizing Feature Map (SOFM)
4. F-Score and Supported Sequential Forward Search (F-SSFS)
5. Decision Tree
6. Fractal Feature Selection

بهینهٔ تعداد ویژگی‌ها کمک می‌کند، استفاده کردند. نتایج نشان از بالا بودن عملکرد این روش نسبت به روش‌های مشابه داشته است (لی، ژی و یا، ۲۰۱۱).

در زمینهٔ پیش‌بینی روند حرکت قیمت سهم در داخل کشور مطالعاتی انجام نشده است، ولی گفتنی است که کاربرد روش ماشین بردار پشتیبان در پیش‌بینی درماندگی مالی شرکت‌ها بررسی شده است (راعی و فلاح‌پور، ۱۳۸۷). در ادامه به مطالعات انجام شده در زمینهٔ پیش‌بینی شاخص و قیمت سهام در داخل کشور اشاره شده است.

یکی از پژوهش‌های مربوط به پیش‌بینی شاخص بورس اوراق بهادار در سال‌های اخیر، به‌روش شبکه‌های عصبی انجام شده است (سینایی، مزتصوی و تیموری اصل، ۱۳۸۴). نتایج پژوهش نشان داده است که شبکه‌های عصبی عملکرد بهتری نسبت به مدل خطی ARIMA برای پیش‌بینی شاخص قیمت دارند.

در پژوهشی (منجمی، ابزری و رعیتی شوازی، ۱۳۸۸)، قیمت سهام در بازار بورس اوراق بهادار، به‌کمک شبکهٔ عصبی فازی و الگوریتم‌های ژنتیک پیش‌بینی شده و نتایج آن با روش شبکهٔ عصبی مصنوعی مقایسه شده است. نتایج نشان داده است که مدل ترکیبی شبکه‌های عصبی فازی و الگوریتم‌های ژنتیک پیش‌بینی‌های بسیار مناسب‌تری داشته و در مقایسه با شبکه‌های عصبی منفرد، از سرعت بالاتر و توانایی تقریب قوی‌تری برای پیش‌بینی قیمت سهام برخوردار است.

کوشا (۱۳۸۸)، به بررسی سودمندی مدل‌های ارزش‌گذاری، در پیش‌بینی قیمت سهام در بورس اوراق بهادار تهران پرداخت و توانایی مدل‌های ارزش‌گذاری سهام، شامل مدل نسبت قیمت به سود (ضرایب تکاثری) و مدل ارزش افزودهٔ بازار در پیش‌بینی قیمت بازار سهام را طی یک دورهٔ چهار ساله (۱۳۸۳-۱۳۸۶) در سه صنعت فلزات اساسی، خودرو و ساخت قطعات و صنعت سیمان بررسی کرد. نتایج حاصل حاکی از آن بود که روش‌های ارزش‌گذاری ضریب قیمت به سود و ارزش افزودهٔ بازار می‌توانند در بررسی و پیش‌بینی قیمت و بازده سهام ایران مفید واقع شوند.

عبادی (۱۳۸۸)، در مطالعهٔ خود به پیش‌بینی شاخص کل قیمت سهام در بازار بورس تهران با استفاده از شبکه‌های عصبی مصنوعی پرداخت و نشان داد، شبکهٔ عصبی در برآوردن شاخص کل قیمت سهام بورس اوراق بهادار تهران از کارایی بالایی برخوردار است.

حسن‌زاده (۱۳۸۸)، قیمت سهام را با استفاده از برنامه‌ریزی ژنتیک پیش‌بینی کرده است. فرضیهٔ نخست او افزایش متوسط خطا در نتیجهٔ افزایش تعداد روزهای پیش‌بینی بود. این فرضیه

در همهٔ موارد تأیید شد. فرضیهٔ دوم، نسبت اجرای بهینه به تعداد روزهای پیش‌بینی را بزرگتر مساوی یک نشان داد. این فرضیه تنها در ۷۵ درصد موارد تأیید شد.

حسینی (۱۳۸۸)، در پایان‌نامهٔ کارشناسی ارشد خود به پیش‌بینی قیمت سهام با استفاده از شبکه‌های عصبی در بازار بورس شیراز پرداخته است. در این پژوهش مدل‌سازی و پیش‌بینی رفتار بازار بورس به‌کمک شبکهٔ عصبی انجام گرفت. نامبرده به‌منزلهٔ مطالعهٔ موردی، سه شرکت را انتخاب کرد. نتایج این پژوهش نشان می‌دهد که شبکهٔ عصبی پیشنهاد شده می‌تواند قیمت بستهٔ روز آیندهٔ هر سه شرکت را با دقت بسیار خوبی پیش‌بینی کند. این پژوهش نشان داد، اگر متغیرهای ورودی را از داده‌های تکنیکال استفاده کنیم، می‌توانیم شرکت‌های مختلف یا حتی صنعت‌های مختلف را تنها با یک شبکه مطالعه کنیم.

خمیسی (۱۳۸۹)، قابلیت پیش‌بینی مدل PEG[1] در مقایسه با مدل P/E برای تعیین قیمت سهام در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را بررسی کرد. خمیسی جامعهٔ آماری پژوهش خود را کلیهٔ شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران تا پایان سال ۱۳۷۷ در نظر گرفت که دورهٔ زمانی ابتدای سال ۱۳۸۱ تا پایان سال ۱۳۸۵ را شامل می‌شد. نتایج نشان داد، نسبت P/E در مقایسه با نسبت PEG طی دورهٔ بررسی از ثبات بیشتری برخوردار بوده است و پیش‌بینی قیمت سهام با استفاده از PEG از دقت بیشتری برخوردار است.

هاشمی (۱۳۸۹)، بررسی تأثیر فاکتورهای رفتاری بر پیش‌بینی قیمت سهام با استفاده از مدل شبکه‌های عصبی رگرسیونی و جلوسو را مورد توجه قرار داد و با مطالعهٔ سهام ده شرکت تشکیل‌دهندهٔ شاخص داجونز به این نتیجه رسید که فاکتورهای رفتاری در پیش‌بینی قیمت سهام نُه شرکت از ده شرکت مؤثر هستند و دقت پیش‌بینی را به طرز چشمگیری افزایش می‌دهند.

اشهر (۱۳۸۹)، در پایان‌نامهٔ کارشناسی ارشد خود در دانشگاه یزد، شاخص سهام را با استفاده از الگوریتم پرواز پرندگان[2] تخمین زد و سپس نتیجهٔ آن را با مدل‌های سنتی رایج در بورس اوراق بهادار تهران مقایسه کرد. نتایج پژوهش حاکی از آن است که الگوریتم پرواز پرندگان نسبت به تمام مدل‌های سنتی، دقیق‌ترین پیش‌بینی را ارائه می‌کند که این مدل هوشمند خطای برآورد را نسبت به بهترین مدل سنتی بهبود می‌بخشد. همچنین، الگوریتم پرواز پرندگان واریانس خطا را نسبت به مدل‌های سنتی بسیار کاهش می‌دهد که این نشان‌دهندهٔ دقت بالای این مدل هوشمند در پیش‌بینی است.

---

1. Price/Eearning-to-Growth
2. Particle Swarm Optimization (PSO)

مدل مفهومی

این پژوهش کاربردی بوده، مبتنی بر پـژوهش‌هـای میـدانی اسـت؛ یعنـی بـر مبنـای اطلاعـات جمع‌آوری‌شده از بازار بورس تهران، فرضیهٔ پژوهش آزمون و سپس نتایج حاصله به کـل جامعـهٔ آماری تعمیم داده می‌شود. در این پژوهش، ابتدا پیش‌بینی روند حرکتی قیمـت سـهام بـه‌کمـک مدل ماشین بردار پشتیبان انجام گرفت و دقت پیش‌بینی این روش محاسـبه شـد. سـپس مـدل ترکیبی ماشین بردار پشتیبان برپایهٔ الگوریتم ژنتیک را طراحی کرده و دقت پیش‌بینی ایـن روش نیز محاسبه شده است. در آخر به مقایسهٔ دقت پیش‌بینی روش ترکیبی با روش‌های ماشین بردار پشتیبان پرداخته شده است.

هدف اصلی این پژوهش، طراحی مدلی است که بتواند روند حرکتی قیمت سهام را برای یک روز آتی پیش‌بینی کند. منظور از روند حرکتی قیمت سهام، مثبت یا منفـی بـودن قیمـت پایـانی سهام در یک روز بعد تعریف شده است. فرض پژوهش بر این پایه استوار اسـت کـه اگـر قیمـت پایانی سهم مورد نظر در روز بعد نسبت به امروز مثبت باشـد، خروجـی مـورد نظـر بـرای امـروز «یک» تعریف شده، اگر قیمت پایانی سهم مورد نظر در روز بعد نسبت به امروز صـفر یـا منفـی باشد و به تعبیری مثبت نباشد، خروجی مورد نظر برای امروز «منفی یـک» تعریـف شـده اسـت. گفتنی است، برای پیش‌بینی روند حرکتی قیمت سهام برای یک روز آینـده، از تمـامی داده‌هـای روز قبل استفاده می‌شود.

مدل‌سازی به‌وسیلهٔ ۴۴ متغیر ورودی انجام گرفت. از مجموع ۴۴ متغیـر، نُـه متغیـر اولیـه یـا اصلی (متغیرهای یک تا نُه که در جدول متغیرها فهرست شده‌انـد) بـرای هـر روز معـاملاتی، از شرکت مدیریت فناوری بورس تهران گردآوری شده است و مابقی متغیرها که شـامل نماگرهـای تکنیکی هستند، با استفاده از فرمول‌های ریاضی به‌دست آمده‌اند و هر نمـاگر بـا توجـه بـه دورهٔ متداول خود محاسبه شده است. در نهایت، برای هر شرکت یک مـاتریس متشـکل از ۴۴ متغیـر ورودی و یک متغیر خروجی که ۱ یا ۱- است، برای هر روز و به تعداد روزهای مورد نظـر بـرای هر شرکت به‌دست آمده است که به‌منزلهٔ ورودی مدل‌ها استفاده می‌شود. معیار مقایسـهٔ عملکـرد مدل‌ها نیز دقت پیش‌بینی درست روند حرکتی قیمت سهام است. بر این اساس که هم پیش‌بینی مثبت و هم پیش‌بینی منفی روز آتی را لحاظ می‌کند و چنانچه پیش‌بینـی درسـت باشـد، هـر دو حالت یادشده، معیار عملکرد محاسبه می‌شود و دقت مدل مورد نظر به‌دست می‌آید.

فرضیهٔ پژوهش به این صورت تعریف شده است: روش ترکیبی پیش‌بینـی بـا ماشـین بـردار پشتیبان برپایهٔ الگوریتم ژنتیـک عملکـرد بهتـری از روش ماشـین بـردار پشـتیبان سـاده بـرای پیش‌بینی روند حرکت قیمت سهم دارد.

در رابطه با قلمرو زمانی پژوهش نیز، به‌دلیل اینکه برای آموزش ماشین بردار پشتیبان به داده‌های زیادی نیاز است، بازهٔ زمانی از ابتدای سال ۱۳۸۳ تا پایان سال ۱۳۸۹ انتخاب شده است.

روش‌شناسی پژوهش

طرح‌ریزی مدل ماشین بردار پشتیبان بر پایهٔ الگوریتم ژنتیک

در این مدل برای تمامی شرکت‌ها، یک مدل SVM طراحی شده است، به‌طوری که تعداد متغیرهای ورودی نهایی را الگوریتم ژنتیک مشخص کرده است. در این مدل تمام ۴۴ متغیر به‌منزلهٔ ورودی عمل می‌کنند، ولی نه برای SVM، بلکه برای الگوریتم ژنتیک. بعد از وارد شدن ۴۴ متغیر به الگوریتم ژنتیک، بر اساس آنچه که پیش‌تر گفته شد، یک تابع برازش لحاظ شده است که در مدل یادشده، الگوریتم هر بار که یک کروموزم از یک نسل را بررسی می‌کند، متغیرهای لحاظ شده به‌وسیلهٔ آن کروموزم را هر تعداد که باشند، با SVM مورد آزمون قرار داده، خطای آن را به‌دست می‌آورد و به‌منزلهٔ برازش به آن کروموزم نسبت می‌دهد. بعد از آزمودن کروموزم‌ها در هر نسل از الگوریتم ژنتیک، بهترین برازش‌ها که برابر با کمترین خطاها هستند، انتخاب شده و به نسل بعد انتقال می‌یابند. در خلال این کار، برای تولید نسل بعدی از روش‌های جهش و تقاطع نیز استفاده شده است. گفتنی است، پارامترهای کنترلی الگوریتم ژنتیک با آزمون و خطا به‌دست آمده است. درنهایت بعد از سپری شدن چندین تکرار و تولید نسل‌های بعدی، بر اساس پارامترهای کنترلی توقف تکرار، الگوریتم ژنتیک متوقف شده، بهترین کروموزم را که دربردارندهٔ تعدادی از متغیرها بوده و تا آن لحظه از تکرار الگوریتم محاسبه شده، مشخص می‌شود. به‌طور کلی در این مدل، SVM به‌کمک الگوریتم ژنتیک، از ۴۴ متغیر ورودی، تعدادی از آنها را که برای پیش‌بینی روند حرکتی سهام مناسب‌تر هستند انتخاب کرده، بدین‌صورت مدل مورد نظر را بهینه‌سازی می‌کند و متغیرهایی را که به‌منزلهٔ سیگنال منفی عمل‌کرده و باعث عملکرد نامناسب مدل پیش‌بینی و کاهش دقت پیش‌بینی درست شده، از فرآیند پیش‌بینی کنار گذاشته می‌شوند.

پس از طراحی مدل‌ها و پیش‌بینی نتایج با دو روش مدل ترکیبی و SVM ساده و محاسبهٔ میزان خطای دو روش، به مقایسهٔ دقت پیش‌بینی این روش‌ها پرداخته شده است. از آنجاکه در پژوهش حاضر از مدل دسته‌بندی استفاده شده، یک مقیاس برای مقایسهٔ دقت پیش‌بینی به‌دست آمد که خطای کل پیش‌بینی نامیده می‌شود.

جامعهٔ آماری این پژوهش، شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران هستند. در این پژوهش از روش نمونه‌گیری حذفی سیستماتیک یا روش غربال (یا همان روش قضاوتی) استفاده شد. برای انتخاب شرکت‌های مورد بررسی دو نکته مد نظر قرار گرفته است: نخست، تعداد روزهای معاملاتی شرکت‌ها زیاد باشد؛ این امر سبب می‌شود که آموزش بهتری به الگوریتم بدهیم. دوم، شرکت‌ها از گروه‌های ناهمگن باشند تا از این طریق بتوان به دید کلی از نحوهٔ پیش‌بینی بازار بورس ایران دست یافت. داده‌های این پژوهش، مربوط به سی شرکت از پنجاه شرکت برتر سه‌ماههٔ دوم سال ۱۳۹۰ هستند که این سی شرکت براساس بیشترین تعداد روزهای کاری رتبه‌بندی شده‌اند.

اسامی این شرکت‌های عبارتند از: بانک اقتصاد نوین، چادرملو، بانک کار آفرین، توسعهٔ صنایع بهشهر، سیمان فارس و خوزستان، فارسیت درود، سایپا آذین، گروه بهمن، الکتریک خودرو شرق، توسعهٔ معادن روی ایران، داروسازی جابربن‌حیان، کالسیمین، کیمیدارو، گل گهر، مس شهید باهنر، پتروشیمی خارک، تکین‌کو، مهرکام‌پارس، نوسازی و ساختمان تهران، تولی‌پرس، شهید قندی، صنایع بهشهر، صنایع آذرآب، فرآوری مواد معدنی، ایران ترانسفو، صنایع لاستیکی سهند، خدمات انفورماتیک، ملی مس، ایران ارقام و سیمان غرب.

گفتنی است، ارزش بازار شرکت‌های نمونهٔ پژوهش، در پایان دوره پژوهش برابر با ۲۴۵,۴۳۹ میلیارد ریال به‌دست آمد. همچنین ارزش کل بازار نیز در همان تاریخ برابر با ۱,۱۱۵,۶۳۶ میلیارد ریال بوده است.

متغیرهای پژوهش

در جدول ۱ فهرستی از متغیرهای استفاده‌شده در پژوهش حاضر آورده شده است.

جدول۱. متغیرهای پژوهش

نام متغیر	تعریف
قیمت آغازین	قیمتی است که در شروع معاملات برای هر ورقهٔ بهادار در تابلو بورس اعلام می‌شود.
قیمت بالا	بالاترین قیمتی است که در یک روز رسمی معاملاتی برای هر ورقهٔ بهادار در تابلو بورس اعلام می‌شود.
قیمت پایین	پایین‌ترین قیمتی است که در یک روز رسمی معاملاتی برای هر ورقهٔ بهادار در تابلو بورس اعلام می‌شود.
قیمت پایانی	قیمت هر ورقهٔ بهادار در پایان هر روز رسمی معاملاتی بورس است که محاسبه شده و «بورس» آن را اعلام می‌کند.

ادامهٔ جدول ۱. متغیرهای پژوهش

نام متغیر	تعریف
قیمت میانگین	میانگینی از قیمت‌های معامله‌شدهٔ هر ورقهٔ بهادار در یک روز رسمی معاملاتی بورس است.
تعداد خریداران	تعداد خریداران در یک روز رسمی معاملاتی برای هر ورقهٔ بهادار در بورس است.
دفعات معامله	تعداد معاملات در یک روز رسمی معاملاتی برای هر ورقهٔ بهادار در بورس است.
تعداد سهام معامله‌شده	تعداد سهام معامله‌شده یا حجم معاملات در یک روز رسمی معاملاتی برای هر ورقهٔ بهادار در بورس است.
قیمت به سود هر سهم	از تقسیم قیمت روز سهم به میزان سود هر سهم به‌دست می‌آید.
میانگین متحرک سادهٔ پنج‌روزه	شاخص‌های پسرو یا پیرو روند هستند که نوسان‌های نامنظم بازار را دو دورهٔ پنج‌روزه تعدیل می‌کنند.
میانگین متحرک سادهٔ بیست‌روزه	شاخص‌های پسرو یا پیرو روند هستند که نوسان‌های نامنظم بازار را در دورهٔ بیست‌روزه تعدیل می‌کنند.
میانگین متحرک موزون	شاخصی است که میانگین قیمت یک سهم را طی یک دورهٔ زمانی با تأکید ویژه بر قیمت‌های جدید، در دورهٔ پنج‌روزه نشان می‌دهد.
میانگین متحرک نمایی پنج‌روزه	در این میانگین آخرین داده ارزش بیشتری دارد و در دورهٔ پنج‌روزه، هیچ داده‌ای به‌طور واقعی از محاسبه خارج نمی‌شود.
میانگین متحرک نمایی بیست‌وشش‌روزه	در این میانگین آخرین داده ارزش بیشتری دارد و در دوره ۲۶ روزه هیچ داده‌ای به‌طور واقعی از محاسبه خارج نمی‌شود.
نوسان نمای مومنتوم چاند	برای به‌دست آوردن اندازهٔ حرکت (مومنتوم) خالص در دورهٔ نه‌روزه استفاده می‌شود.
مومنتوم	سرعت تغییر قیمت را در دورهٔ پنج‌روزه اندازه‌گیری می‌کند.
نرخ تغییر قیمت	یک شاخص مومنتوم است که سرعت تغییر قیمت را در دورهٔ پنج‌روزه اندازه‌گیری می‌کند.
شاخص قدرت نسبی	افزایش و کاهش را در قیمت‌های پایانی برای یک دورهٔ معین چهارده‌روزه اندازه‌گیری می‌کند.
استوکاستیک K%	موقعیت‌های اشباع خرید و اشباع فروش را در دورهٔ پنج‌روزه تعیین می‌کند.
استوکاستیک D%	موقعیت‌های اشباع خرید و اشباع فروش را بر اساس K% در دورهٔ پنج‌روزه تعیین می‌کند.

ادامهٔ جدول ۱. متغیرهای پژوهش

نام متغیر	تعریف
میانگین متحرک همگرا / واگرا	یک شاخص پیرو روند است؛ به این معنی که در تشخیص روند فعلی سهم (صعودی یا نزولی) کمک می‌کند.
شاخص پراکندگی	رابطهٔ قیمت بسته‌شدن و میانگین متحرک ساده را در دورهٔ پنج‌روزه نشان می‌دهد.
نوسان نمای قیمت	شاخصی است که تفاوت دو میانگین متحرک را نشان می‌دهد.
شاخص کیو استیک	میانگین متحرک پنج‌روزه تفاوت میان قیمت‌های باز شدن و قیمت‌های بسته شدن یک سهم است.
فیلتر افقی عمودی	برای تشخیص بازارهای رونددار از بازارهای بدون روند در دورهٔ پنج‌روزه استفاده شده است.
شاخص اندازهٔ حرکت دینامیک	تقریباً مشابه RSI است با این تفاوت که تعداد دوره‌های زمانی در DMI متغیر است.
شاخص اندازهٔ حرکت دورن روزانه	ترکیبی از RSI و تحلیل شمعی است که در دورهٔ ده‌روزه محاسبه شده است.
نوسان‌نمای ویلیامز	یک شاخص مومنتوم است که شرایط اشباع خرید و اشباع فروش را در دورهٔ چهارده‌روزه نشان می‌دهد.
میانگین نزول	یک میانگین متحرک از کاهش میان قیمت‌های متوالی در دورهٔ چهارده‌روزه است.
نوسان‌نمای غیر روندی قیمت	با کم‌کردن یک میانگین متحرک از قیمت، روند را در قیمت‌ها در دورهٔ چهارده‌روزه حذف می‌کند.
شاخص تسهیل بازار	این شاخص برای افزایش دقت سیگنال‌ها، هم قیمت و هم حجم را در محاسباتش دخیل می‌کند.
شاخص روند قیمت و حجم	حاصل تجمعی از حجم است که بر اساس تغییرات نسبی قیمت‌های پایانی تنظیم شده است.
شاخص حجم منفی	برای شناسایی بازارهای صعودی و نزولی به‌کار می‌رود.
شاخص حجم مثبت	برای شناسایی بازارهای صعودی و نزولی به‌کار می‌رود.
شاخص حجم تعادلی	برای تأیید روند جاری و نیز، اخطار نسبت به تغییر احتمالی روند استفاده می‌شود.
شاخص قیمت نمونه	گونهٔ دیگری از قیمت میانگین برای هر دوره که با یک خط ساده طراحی شده است.

ادامهٔ جدول ۱. متغیرهای پژوهش

نام متغیر	تعریف
نوارهای بولینگر میانی	نوار میانی خطوط بولینگر است که به موازات یک میانگین متحرک چهارده‌روزه قرار گرفته است.
نوارهای بولینگر بالا	نوار بالایی خطوط بولینگر است که به موازات یک میانگین متحرک چهارده‌روزه قرار گرفته است.
نوارهای بولینگر پایین	نوار بالایی خطوط بولینگر است که به موازات یک میانگین متحرک چهارده‌روزه قرار گرفته است.
شاخص تجمع و توزیع ویلیامز	خط متحرکی است که از مجموع مقادیر مثبت تجمع (تقاضا) و مقادیر منفی توزیع (عرضه) تشکیل شده است.
شاخص قیمت پایانی موزون	شاخص قیمت پایانی موزون مشابه شاخص قیمت نوعی است، ولی ارزش بیشتری برای قیمت پایانی قائل است.
شاخص تجمع و توزیع	بر پایهٔ رابطهٔ حجم با قیمت عمل می‌کند.
شاخص نوسان‌نمای چاکین	یک مومنتومی از خط تجمع و توزیع است.
شاخص نوسانات چاکین	شاخصی از نرخ تغییرات محدودهٔ معاملات است.

یافته‌های پژوهش

برای اینکه درستی دقت پیش‌بینی دو مدل مقایسه شود، داده‌های تمامی شرکت‌ها به پنج دوره تقسیم‌شده و نتایج حاصل از پیش‌بینی مدل‌ها برای تمام این دوره‌ها محاسبه شده است. برای ارزیابی همه‌جانبهٔ توانمندی دو مدل، باید مقدار خطای نوع اول و دوم محاسبه شود که این کار برای هر یک از روش‌ها به‌طور جداگانه انجام گرفت و نتایج این محاسبات در جدول‌های جداگانه آمده است. گفتنی است، خطای نوع اول، خطاهایی هستند که خروجی مدل در واقعیت مثبت بوده و در پیش‌بینی هم مثبت برآورد شده است یا خروجی در واقعیت منفی بوده، در پیش‌بینی هم منفی برآورد شده است. خطای نوع دوم، خطاهایی هستند که خروجی مدل در واقعیت مثبت بوده، در پیش‌بینی منفی برآورد شده است یا خروجی در واقعیت منفی بوده، در پیش‌بینی مثبت برآورد شده است. این نسبت‌ها، خروجی ماتریس دقت[1] است که در این قسمت به‌صورت خلاصه ارائه خواهد شد. نتیجهٔ نهایی این ماتریس از ضرب درصد داده‌های مثبتی که مثبت پیش‌بینی شده‌اند و درصد داده‌های منفی که منفی پیش‌بینی شده‌اند به‌دست می‌آید.

1. Confusion Matrix

جدول ۲. نتایج عددی ماتریس دقت برای دو مدل ترکیبی و SVM ساده

		مدل ترکیبی		پیش‌بینی		مدل SVM	
				مثبت	منفی		
دورهٔ اول	واقعیت	مثبت	۷۷۷	۶۱۵	۵۳۳	مثبت	واقعیت
		منفی	۳۲۳	۲۵۵	۱۷۷۸	منفی	
دورهٔ دوم	واقعیت	مثبت	۷۰۵	۴۹۹	۵۱۰	مثبت	واقعیت
		منفی	۲۹۵	۲۰۸	۲۰۰۳	منفی	
دورهٔ سوم	واقعیت	مثبت	۸۱۵	۶۹۷	۵۱۵	مثبت	واقعیت
		منفی	۲۷۴	۲۹۲	۱۸۲۳	منفی	
دورهٔ چهارم	واقعیت	مثبت	۷۶۹	۷۱۵	۵۸۶	مثبت	واقعیت
		منفی	۳۳۹	۳۵۱	۱۶۶۸	منفی	
دورهٔ پنجم	واقعیت	مثبت	۸۲۷	۶۷۳	۶۵۲	مثبت	واقعیت
		منفی	۴۳۵	۳۴۳	۱۵۲۰	منفی	

جدول ۳. نتایج درصدی ماتریس دقت برای دو مدل ترکیبی و SVM ساده

		مدل ترکیبی		پیش‌بینی		مدل SVM	
				مثبت	منفی		
دورهٔ اول	واقعیت	مثبت	۰/۵۹	۰/۴۵	۰/۴۱	مثبت	واقعیت
		منفی	۰/۱۵	۰/۱۲	۰/۸۵	منفی	
دورهٔ دوم	واقعیت	مثبت	۰/۵۸	۰/۴۰	۰/۴۲	مثبت	واقعیت
		منفی	۰/۱۳	۰/۱۰	۰/۸۷	منفی	
دورهٔ سوم	واقعیت	مثبت	۰/۶۱	۰/۵۰	۰/۳۹	مثبت	واقعیت
		منفی	۰/۱۳	۰/۱۴	۰/۸۷	منفی	
دورهٔ چهارم	واقعیت	مثبت	۰/۵۷	۰/۵۱	۰/۴۳	مثبت	واقعیت
		منفی	۰/۱۷	۰/۱۶	۰/۸۳	منفی	
دورهٔ پنجم	واقعیت	مثبت	۰/۵۶	۰/۴۶	۰/۴۴	مثبت	واقعیت
		منفی	۰/۲۲	۰/۱۸	۰/۷۸	منفی	

جداول بالا از نتایج دورهٔ اول تا پنجم، برای هر دو مدل به‌دست آمده است. داده‌های قید شده در خانه‌های جدول ۲، مجموع نتایج ماتریس دقت برای هر سی شرکت است؛ در جدول ۳ نیز نسبت درصدی آنها ارائه شده که به‌صورت کلی برای هر دوره محاسبه شده است. همان‌طور

که ملاحظه می‌شود در هر پنج دوره، تعداد داده‌هایی که مثبت یا منفی بوده و درست پیش‌بینی شده‌اند، در مدل ترکیبی بیشتر از مدل SVM است که برتری این مدل نسبت به مدل SVM را نشان می‌دهد. بر اساس نتایج بالا؛ یعنی برتری مدل ترکیبی، مقایسه‌ای برای ارزیابی دقت کل پیش‌بینی بین دو مدل مورد استفاده در این پژوهش انجام شده است (جدول ۴).

جدول ٤. نتایج دقت کل پیش‌بینی برای دو مدل ترکیبی و SVM ساده

دقت کل پیش‌بینی (درصد)		
دوره	مدل ترکیبی SVM با الگوریتم ژنتیک	مدل SVM
۱	۷۵/۵۸	۷۰/۴۳
۲	۷۷/۰۷	۷۱/۴۹
۳	۷۶/۹۲	۷۱/۳۲
۴	۷۳/۵۷	۶۹/۲۶
۵	۶۹/۱۲	۶۵/۴
میانگین	۷۴/۴۵	۶۹/۵۸

جدول ۴ نتایج کلی به‌دست آمده از سی شرکت را برای مدل‌های ترکیبی و SVM به تفکیک دوره نشان می‌دهد. اعداد قید شده در جدول میانگینی از نتایج پیش‌بینی مدل‌ها برای سی شرکت در دورهٔ مورد نظر هستند. همان‌طور که از نتایج جدول مشخص است، دقت کل پیش‌بینی در مدل ترکیبی نسبت به مدل SVM در هر پنج دوره بالاتر به‌دست آمده که نشان از برتری مدل ترکیبی در مقابل مدل SVM دارد. برای بررسی فرضیهٔ پژوهش، آزمون ویلکاکسون حاکی از معناداری تفاوت دقت پیش‌بینی دو روش بود و بالاتر بودن دقت پیش‌بینی مدل ترکیبی را نشان داد. در این آزمون آمارهٔ Z برابر با ۱۰/۶۲۴- و آمارهٔ P برابر با ۰/۰۰۰ به‌دست آمد.

جدول ۵. نتایج آزمون ویلکاکسون برای فرضیهٔ پژوهش

مدل ترکیبی SVM با ژنتیک در مقابل مدل SVM	
آماره Z	۱۰/۶۲۴-
آماره P	۰/۰۰۰

نتیجه‌گیری و پیشنهادها

مقایسهٔ خطای کل پیش‌بینی برای هر دو روش SVM و GA-SVM، نشان داد مدل ترکیبی GA-SVM قوی‌تر از مدل SVM عمل کرده است و پیش‌بینی را با خطای کمتری انجام داده

است. نتایج حاکی از قابلیت پیش‌بینی مدل ترکیبی برای روزهای منفی بازار بود. به‌طورکلی، دقت پیش‌بینی مدل ترکیبی در مقایسه با مدل SVM ساده، پنج درصد بالاتر به‌دست آمد. نتایج دیگری نیز از پژوهش به‌دست آمد که برگرفته از الگوریتم ژنتیک بود. در حالیکه مدل با ۴۴ ورودی به پیش‌بینی می‌پرداخت، الگوریتم ژنتیک نشان داد که برخی از نماگرها استفاده بیشتری داشته و پرکاربردتر از نماگرهای دیگر بوده‌اند. بر همین اساس نماگر R% ویلیامز با ۲۷ بار تکرار در بین داده‌های مورد بررسی، بیشترین تکرار را داشته است و بعد از آن نوسان نمای مومنتوم چاند، استوکاستیک K% و D% با ۲۴ بار در رده‌های بعدی قرار گرفتند. گفتنی است، متغیر ورودی قیمت بالا با یازده بار تکرار، کمترین تکرار را در بین ۴۴ متغیر داشته است.

با توجه به اینکه نتایج پژوهش حاضر برتری روش ترکیبی را نسبت به SVM ساده نشان داد، پس می‌توان استفاده از GA-SVM را به جای SVM توصیه کرد. از آنجا که در بسیاری از پژوهش‌ها، از مدل SVM در ترکیب با روش‌های دیگر استفاده شده و برتری روش ترکیبی نسبت به SVM ساده به تأیید رسیده، استفاده از روش ترکیبی با الگوریتم‌های دیگر برای بهینه‌سازی مدل و نیز ترکیب با مدل‌های شبکهٔ عصبی و عصبی فازی، می‌تواند نتایج بهتری را در پیش داشته باشد.

پیشنهاد می‌شود، در بررسی و تحلیل تکنیکال سهام از نماگرهایی استفاده شود که برای هر شرکت در نمونهٔ آماری پژوهش با مدل ترکیبی به‌دست آمد.

پیشنهاد می‌شود، در بررسی تکنیکال سهام از نماگر R% ویلیامز بیشتر استفاده شود.

پیشنهاد می‌شود برای دست یافتن به نتیجهٔ بهتر، مدل‌ها برای داده‌هایی با روزهای معاملاتی بالا به‌کار گرفته شوند.

همچنین در پژوهش‌های آتی می‌توان ترکیب SVM با الگوریتم‌های بهینه‌سازی دیگر، مانند کلونی مورچگان، کلونی زنبورها، روش بهینه‌سازی ازدحام ذرات و... را بررسی کرد. نحوهٔ انتخاب نماگرهای تحلیل تکنیکال، تعداد آنها و کاربرد نماگرها در بازار سرمایهٔ ایران بررسی شود و درنهایت، به‌همراه نماگرهای تحلیل تکنیکال از متغیرهای دیگری چون شاخص‌های کلان اقتصادی بهره جسته و معناداری مدل‌ها با متغیرهای مختلف سنجیده شود.

منابع

اشهر، م. (۱۳۸۹). *پیش‌بینی شاخص سهام با استفاده از الگوریتم پرواز پرندگان و مقایسه آن با مدل‌های سنتی رایج در بورس اوراق بهادار تهران*. پایان‌نامهٔ کارشناسی ارشد، دانشکدهٔ مدیریت و اقتصاد دانشگاه یزد.

البرزی، م. (۱۳۸۸). *الگوریتم ژنتیک*. تهران: انتشارات دانشگاه صنعتی شریف.

حسن‌زاده، س. (۱۳۸۸). *پیش‌بینی قیمت سهام با استفاده از برنامه‌ریزی ژنتیک*. پایان‌نامهٔ کارشناسی ارشد، دانشکدهٔ مدیریت و اقتصاد دانشگاه تربیت مدرس، تهران.

حسینی، م. (۱۳۸۸). پیش‌بینی *قیمت سهام با استفاده از شبکه‌های عصبی (مطالعهٔ موردی: بازار بورس شیراز)*. پایان‌نامهٔ کارشناسی ارشد، دانشگاه شیراز.

خمیسی، ح. (۱۳۸۸). *بررسی قابلیت پیش‌بینی مدل PEG در مقایسه با مدل P/E برای تعیین قیمت سهام در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران*. پایان‌نامهٔ کارشناسی ارشد، دانشکدهٔ علوم اجتماعی و اقتصادی دانشگاه الزهرا، تهران.

راعی، ر. و فلاح‌پور، س. (۱۳۸۷). کاربرد ماشین بردار پشتیبان در پشی‌بینی درماندگی مالی شرکت‌ها با استفاده از نسبت‌های مالی. *بررسی‌های حسابداری و حسابرسی*، ۱۵(۵۳): ۳۴– ۱۷.

سینایی، ح.؛ مرتضوی، س. و تیموری اصل، ی. (۱۳۸۴). پیش‌بینی شاخص بورس اوراق بهادار با استفاده از شبکه‌های عصبی مصنوعی. *بررسی‌های حسابداری و حسابرسی*، ۱۲(۴۱): ۸۳–۵۹.

شیرمحمدی، س. (۱۳۸۹). *پیش‌بینی قیمت سهام در بورس اوراق بهادار تهران با استفاده از مدل ترکیبی ARIMA و رگرسیون فازی در فاصلهٔ زمانی (۱۳۸۸–۱۳۸۳) و مقایسهٔ کارایی آن با مدل رگرسیون و ARIMA ساده*. پایان‌نامهٔ کارشناسی ارشد، دانشکده علوم اجتماعی و اقتصادی دانشگاه الزهرا، تهران.

عبادی، ا. (۱۳۸۸). *پیش‌بینی شاخص کل قیمت سهام در بازار بورس تهران با استفاده از شبکه‌های عصبی مصنوعی*. پایان‌نامهٔ کارشناسی ارشد. دانشکدهٔ اقتصاد و علوم اجتماعی دانشگاه بوعلی سینا، همدان.

عبده تبریزی، ح. و جوهری، ه. (۱۳۷۵). بررسی کارآمدی شاخص بورس اوراق بهادار تهران. *تحقیقات مالی*، ۳ (۲): ۶۱–۴۷.

کوشا، س. (۱۳۸۸). *بررسی سودمندی مدل‌های ارزشگذاری در پیش‌بینی قیمت سهام در بورس اوراق بهادار تهران*. پایان‌نامهٔ کارشناسی ارشد، دانشکدهٔ مدیریت و اقتصاد دانشگاه تربیت مدرس، تهران.

هاشمی، ا. (۱۳۸۹). *تأثیر فاکتورهای رفتاری بر پیش‌بینی قیمت سهام با استفاده از مدل شبکه‌های عصبی رگرسیونی و جلوسو (مورد مطالعه: سهام ده شرکت تشکیل دهندهٔ شاخص داجونز*. پایان‌نامهٔ کارشناسی ارشد، دانشکدهٔ صنایع دانشگاه علم و فرهنگ، تهران.

منجمی، س.ا.؛ ابزری، م.؛ رعیتی شوازی، ع. (۱۳۸۸). پیش‌بینی قیمت سهام در بازار بورس اوراق بهادار با استفاده از شبکهٔ عصبی فازی و الگوریتم‌های ژنتیک و مقایسهٔ آن با شبکهٔ عصبی مصنوعی. *فصلنامهٔ اقتصاد مقداری*، ۶(۳): ۲۶-۱.

یاری، س. و خانلو، ب. (۱۳۸۷). *طبقه‌بندی و خوشه‌بندی با استفاده از SVM*. پروژهٔ کارشناسی رشتهٔ علوم کامپیوتر. دانشکدهٔ ریاضی آمار و علوم کامپیوتر دانشگاه تهران، تهران.

Abe, S. (2005). *Support Vector Machines for Pattern Classifications, Springer*. London.

Nair, B.B., Mohandas, V.P. & Sakthivel, N.R. (2010). A Genetic Algorithm Optimized Decision Tree-SVM based Stock Market Trend Prediction System. *(IJCSE) International Journal on Computer Science and Engineering*, 2 (9): 2981-2988.

Huang, C.L. & Tsai, Ch.Y. (2009). A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. *Expert Systems with Applications*. 36 (2): 1529-1539.

Chong, E. K.P. & Stanislaw, Z.H. (2001). *An Introduction to Optimization*. Wiley and Sons, New York.

Choudry, R. & Grag, K. (2008). A Hybrid Machine Learning System for Stock Market Forecasting. *World Academy of Science, Engineering and Technology*, 39. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.193.3052&rep=rep1&type=pdf.

Cristianini, N. & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernal-Based Learning Methods*. Cambridge University Press, Cambridge, UK.

Goldberg, D.E. (1989). *Genetic Algorithms in Search Optimization and Machine Learning*. Addison-Wesley.

Gunn, S.R. (1998). *Support Vector Machines for Classifications and Regression*. Technical Report, School of Electronics and Computer.

Ni, L.P., Ni, Zh. W., & Gao, Y.Zh. (2011). Stock trend prediction based on fractal feature selection and support vector machine. *Expert Systems with Applications*, 38(5): 5569-5576.

Lee, M.Ch. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, 36(8): 10896-10904.

Hsu, Sh., Hsieh, JJ.P.A., Chih, T.Ch. & Hsu, K.Ch. (2009). A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Systems with Applications*, 36(4): 7947-7951.

Huang, Sh-Ch. & Wu, T.-K. (2008). Integrating GA-based time-scale feature extractions with SVMs for stock index forecasting. *Expert Systems with Applications*, 35(4): 2080-2088.

Kara, Y., Boyacioglu, M.A. & Baykan, K. Ö (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, 38(5): 5311-5319.

Yim, J. (2002). A Comparison of Neural Networks with Time Series Models for Forecasting Returns on a Stock Market Index. *Lecture Notes in Computer Science*, 2358: 25-35.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.

Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley & Sons, NY.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی