فصلنامه علمی پیشرفت‌های مالی و سرمایه‌گذاری
سال دوم، زمستان ۱۴۰۰– شماره ۵
صفحات ۲۲–۱
نوع مقاله: پژوهشی

پیش‌بینی قیمت سهام در بازار مالی با استفاده از الگوریتم‌های ترکیبی GA-SVM

امید مهدی عبادتی[۱]
محمدعلی جعفری[۲]
نسیم داودی فر[۳]

تاریخ دریافت: ۱۴۰۰/۰۷/۲۲
تاریخ پذیرش: ۱۴۰۰/۱۲/۰۱
شناسه دیجیتال: 10.30495/afi.2022.1945089.1063

چکیده

هدف مقاله حاضر پیش‌بینی قیمت سهام به‌وسیله الگوریتم ترکیبی GA-SV می‌باشد. پیش‌بینی سری‌های زمانی مانند پیش‌بینی قیمت سهام یکی از مهم‌ترین مشکلات در حوزه مالی است، زیرا داده‌ها ناپایدار بوده و دارای متغیرهای نویزی می‌باشند که تحت‌تأثیر عوامل زیادی قرار دارند. در شرایط واقعی نیز شناسایی حرکات سری زمانی شاخص قیمت سهام بسیار پیچیده می‌باشد؛ بنابراین استفاده از یک مدل کلاسیک به‌تنهایی نمی‌تواند پیش‌بینی دقیقی از شاخص‌های قیمت سهام داشته باشد. ازاین‌رو با به‌کارگیری روش‌های ترکیبی می‌توان عدم اطمینان در پیش‌بینی را کاهش داد. در پیش‌بینی قیمت سهام در بخش مالی بیش از ۱۰۰ شاخص برای درک رفتار بازار سهام ایجاد شده است؛ بنابراین شناسایی شاخص‌های مناسب یک مشکل چالش‌برانگیز است. یکی از تکنیک‌هایی که اخیراً برای پیش‌بینی سریال مورد بررسی قرار گرفته است، رگرسیون پشتیبانی بردار (SVR) است. این مطالعه از الگوریتم ترکیبی GA-SVM برای پیش‌بینی شاخص قیمت سهام استفاده می‌کند. نتایج تجربی نشان می‌دهد که الگوریتم پیشنهادی جایگزین مناسب‌تر و امیدوار کننده‌تری برای پیش‌بینی بازار سهام فراهم می‌آورد.

کلید واژه‌ها: پیش‌بینی، قیمت سهام، الگوریتم GA، الگوریتم SVM، بازارهای مالی بین‌الملل.
طبقه‌بندی موضوعی: C6, D4, F3, G2, G4.

---

۱. گروه مدیریت عملیات و فناوری اطلاعات، دانشگاه خوارزمی، تهران، ایران. (نویسنده مسئول). ebadati@khu.ac.ir
۲. گروه ریاضیات مالی، دانشگاه خوارزمی، تهران، ایران.
۳. گروه ریاضیات مالی، دانشگاه خوارزمی، تهران، ایران.

۱- مقدمه

یکی از عناصر کلیدی تصمیم‌گیری‌های مدیریتی، پیش‌بینی می‌باشد که پیشامدهای آینده را با هدف کاهش ریسک تخمین می‌زند. بعلاوه، پیش‌بینی یکی از مهم‌ترین اهداف علوم مالی و اقتصادی است. معمولا پیش‌بینی‌ها صحیح نبوده و دارای خطا است. به همین علت مهم‌ترین عامل مؤثر در انتخاب روش پیش‌بینی، دقت پیش‌بینی است. در چنین شرایطی، داشتن یک مدل پیش‌بینی با دقت بالا به‌منظور حفظ حضور سرمایه‌گذاران به‌عنوان رکن اصلی بازار سرمایه، مورد علاقه تمام سرمایه‌گذاران می‌باشد. پیش‌بینی، یکی از اهداف مدل‌های سری زمانی است. زمانی می‌توان از مدل‌های سری زمانی برای پیش‌بینی استفاده کرد که مدل ساختاری مناسبی وجود نداشته باشد. به عبارت دیگر پیش‌بینی به‌وسیله مدل‌های سری زمانی، در شرایطی که متغیرهای توضیحی به دلیل تعدد بسیار زیاد یا عدم دسترسی به مقادیر عددی آن‌ها، امکان استفاده از آنها وجود ندارد، شیوه مناسبی می‌باشد (دولو و صفری، ۱۳۹۵). با ظهور رایانه دیجیتالی، پیش‌بینی بازار سهام به حوزه فناوری اطلاعات منتقل شده است. با رشد فناوری اطلاعات و هوش مصنوعی و همچنین تأثیرگذاری سری زمانی مالی، پیش‌بینی به یکی از چالش برانگیزترین برنامه ها تبدیل شده است. روند پیش‌بینی آینده، مبتنی بر داده‌های گذشته و موجود است. پیش‌بینی قیمت سهام روشی است که با استفاده از آن داده‌های بورس سهام در زمان مشخص در گذشته برای پیش‌بینی قیمت سهام در دوره آینده پردازش می‌شوند (عبادتی و مرتضوی[1]، ۲۰۱۸).

داده‌های بورس اوراق بهادار حاوی اطلاعات بسیار مفیدی است و برای صحت پیش‌بینی خوب باید حجم زیادی از اطلاعات در نظر گرفته شود. دقت بالاتر دارای نتایج با حداقل خطا است و می‌تواند بر اساس داده‌های گذشته، دوره‌های طولانی را پیش‌بینی کند. در پژوهش‌های اخیر، روند استفاده از هوش مصنوعی به جای روش‌های سنتی وجود دارد. شبکه‌های عصبی مبتنی بر مغز انسان با عملکردهای بسیار برجسته مدل‌سازی می‌شوند و باتوجه‌به قدرت این مدل‌ها می‌توان جهت برخی از مشکلات اقتصادی، راه حل‌هایی را پیدا نمود (دایی و همکاران، ۱۳۹۸). با توسعه اقتصاد ملی و بهبود خدمات مالی، بازار مالی توجه دانشمندان و سرمایه‌گذاران داخلی و خارجی را به خود جلب کرده است. آن‌ها بطور منظم در تلاش برای پیش‌بینی روند بازار، تئوری‌های مختلفی را ارائه می‌دهند که می‌تواند در زمینه پیش‌بینی روند سهام کاربرد داشته باشد. با این حال، از آنجا که بازار تحت تأثیر سیاست‌های ملی، اقتصاد جهانی و منطقه‌ای و همچنین عوامل روانشناختی، انسانی و سایر عوامل است،

1. Ebadati & Mortazavi

پیش‌بینی‌های بازار مالی تمایل به دستیابی به نتایج مطلوب به طور مکرر دارد (پانگ و همکاران[1]، ۲۰۱۸). پیش‌بینی‌کنندگان مالی و معامله‌گران باید خطای پیش‌بینی مسیر را به حداقل برسانند تا انحراف تخمین‌ها را از ارزش‌های واقعی مشاهده شده به حداقل برسانند. با این حال، رفتار در بورس سهام در بسیاری از افراد به عوامل مؤثر کیفی مانند عوامل هیدرولیکی، اقتصادی و طبیعی بستگی دارد. بازار سهام از تنوع پویا و گسترده‌ای برخوردار است (کیو و سانگ[2]، ۲۰۱۶). پیش‌بینی سهام از طریق تجزیه‌وتحلیل داده‌های بازار موضوعی جذاب است. قیمت سهام و مقالات خبری در فرآیندهای پیش‌بینی به کار رفته اند. با این حال، چگونه ترکیب شاخص‌های فنی از قیمت سهام و احساسات خبری از مقالات خبری متنی و اینکه مدل پیش‌بینی قادر به یادگیری اطلاعات متوالی در مجموعه‌های زمانی به صورت هوشمند باشد، هنوز یک مشکل حل نشده است (ونگ و همکاران[3] ، ۲۰۱۷).

فرضیه بازار کارا معتقد است که پیش‌بینی ارزش سهام و اینکه سهام به طور تصادفی رفتار می‌کنند غیرممکن است، اما تجزیه‌وتحلیل تکنیکی اخیر نشان می‌دهد که بیشتر ارزش سهام منعکس‌کننده سوابق قبلی است؛ بنابراین روند حرکت برای پیش‌بینی ارزش‌ها به طور مؤثر ضروری است. در واقع، ارزش سهام به‌طورکلی پویا، غیرپارامتری و غیرخطی است، ازاین‌رو پیش‌بینی قیمت سهام با مشکل ذاتی برخورد می‌کند. علاوه بر این، پیش‌بینی قیمت سهام به دلیل تعداد متغیرهایی که در آن دخیل است، یک مسئله چالش‌برانگیز است (نبی پور و همکاران[4]، ۲۰۲۰). پیش‌بینی قیمت سهام برای سرمایه‌گذاران مفید است تا سود حاصل از معاملات سهام را افزایش دهند. برای پیش‌بینی دو رویکرد سنتی وجود دارد: تحلیل تکنیکی و بنیادی. تجزیه‌وتحلیل فنی می‌کوشد تا برخی از الگوهای تحلیلی بنیادی تاریخ داده را که بر اقتصاد کلی، وضعیت مالی شرکت و مدیریت آن متمرکز شده است، شناسایی کند. به دلیل انفجار اطلاعات در سال های اخیر، حجم زیادی از داده‌ها به صورت آنلاین در دسترس بوده و تجزیه‌وتحلیل دستی این داده‌ها امکان پذیر نیست. تکنیک‌های مختلف یادگیری ماشینی تحت نظارت و بدون نظارت برای ایجاد خودکار جمع‌آوری و تجزیه‌وتحلیل داده ها و پیش‌بینی‌ها ایجاد شده است (ژانگ و همکاران[5]، ۲۰۱۹).

برای کاهش ریسک انتخاب و استفاده از مدل نامناسب، می‌توان از طریق ترکیب روش‌های مختلف پیش‌بینی استفاده نمود، تا بر اساس آن به‌دقت پیش‌بینی بهتری به دست آوریم. استفاده از مدل‌های ترکیبی برای پیش‌بینی قیمت سهام، مانند ایده متنوع‌سازی سبد سهام برای کاهش ریسک می‌باشد که

1. Pang *et al.*
2. Qiu & song
3. Weng *et al.*
4. Nabipour *et al.*
5. Zhang *et al.*

در این حوزه ترکیب پیش‌بینی‌ها همان سبد سهام می‌باشد و منابع ریسک، بازتاب‌دهنده اطلاعات ناقص درباره متغیر هدف و انتخاب مدل نامناسب به دلیل مشخص نبودن فرآیند تولید داده است (جنا و پادهی[1]، ۲۰۱۴). روش‌های یادگیری ماشینی توسط چندین محقق برای پیش‌بینی موفقیت‌آمیز قیمت ابزارهای مالی از داده‌های سری زمانی مالی بازارهای مختلف مورداستفاده قرار می‌گیرند. ماشین‌های بردار پشتیبان SVM[2] روش‌های امیدوارکننده برای پیش‌بینی سری‌های زمانی مالی هستند، زیرا از یک تابع ریسک متشکل از خطای تجربی و یک عبارت تنظیم‌شده که برگرفته از اصل کمینه‌سازی ریسک ساختاری است استفاده می‌کنند. برای بهبود دقت پیش‌بینی، فرد اول باید شاخص‌های مهم را شناسایی کند. برای پیش‌بینی قیمت سهام در بخش مالی بیش از ۱۰۰ شاخص برای درک رفتار بازار سهام توسعه داده شده است (جنا و پادهی، ۲۰۱۴). وجه تمایز پژوهش حاضر با پژوهش‌های دیگر استفاده از مناسب‌ترین شاخص‌های عملکرد برای یافتن بهترین جواب و همچنین یافتن جمعیت بهینه، از کل داده‌های در دسترس، به‌وسیله الگوریتم GA[3] و انتقال آنها به الگوریتم ماشین بردار پشتیبان جهت پیش‌بینی دقیق‌تر قیمت سهام که در مقایسه با دیگر روش‌های پیش‌بینی قیمت سهام دارای خطای کمتری می‌باشد.

در این پژوهش پس از بخش مقدمه، ساختار کلی پژوهش و اهمیت مسائل پیش‌بینی قیمت سهام ارائه خواهد شد. پس از آن مبانی نظری و پیشینه پژوهش ارائه خواهد گردید. سپس به روش پژوهش و پیاده‌سازی الگوریتم‌های پیشنهادی پرداخته خواهد شد. در ادامه به تجزیه‌وتحلیل داده‌ها و آزمون فرضیه‌ها و نهایتاً در بخش آخر به نتیجه‌گیری و بحث پرداخته خواهد گردید.

در این پژوهش از شاخص‌های عملکرد پیش‌بینی با استفاده از معیارهای آماری، یعنی میانگین خطای مربع نرمال شده (NMSE[4])، میانگین خطای مطلق (MAE[5]) و تقارن جهت (DS[6]) ارزیابی می‌شود. همان‌طور که ماهیت بازارها در مناطق مختلف متفاوت است، همچنین از الگوریتم GA برای یافتن جمعیت بهینه از داده‌ها و از الگوریتم SVM برای پیش‌بینی شاخص مناسب قیمت سهام استفاده خواهیم کرد. بعلاوه دو تکنیک یادگیری ماشین: تکنیک ماشین بردار پشتیبان (SVM) و تکنیک ماشین بردار پشتیبان با الگوریتم ژنتیک (GA-SVM) برای پیش‌بینی قیمت‌های آتی ۸ شرکت بین‌المللی مورداستفاده قرار گرفته‌اند. عملکرد این تکنیک‌ها در مقایسه با SVM می‌تواند نتایج عملکرد بهتری را

1. Jena & Padhy
2. Support Vector Machine (SVM)
3. Genetic Algorithm
4. Normalized Mean Square Error
5. Mean Absolute Error
6. Directional Symmetry

در مقایسه با تکنیک SVM فراهم کند. در بهینه‌سازی کار ما از کاربرد GA-SVM در تشخیص و حذف ورودی‌های لازم برای پیش‌بینی یا اجتناب از آسیب‌پذیری مناسب که معمولاً در مدل‌ها با پارامترهای بسیار زیادی اتفاق می‌افتد، حاصل می‌شود. این امر منجر به بهبود سرعت پردازش SVM و افزایش دقت پیش‌بینی شده است.

پیش‌بینی تغییر قیمت سهام به‌عنوان یک فعالیت چالش‌انگیز در پیش‌بینی سری‌های زمانی مالی در نظر گرفته می‌شود. یک پیش‌بینی صحیح از تغییر قیمت سهام می‌تواند سود زیادی را برای سرمایه‌گذاران به بار آورد. توسعه مدل‌های کارآمد برای پیش‌بینی، باتوجه‌به پیچیدگی داده‌های قیمت سهام بسیار دشوار است. پیش‌بینی سری‌های زمانی مانند پیش‌بینی قیمت سهام یکی از مهم‌ترین مشکلات در حوزه مالی است، زیرا داده‌ها ناپایدار بوده و دارای متغیرهای نویزی می‌باشند که تحت‌تأثیر عوامل زیادی قرار دارند. در شرایط واقعی نیز شناسایی حرکات سری زمانی شاخص قیمت سهام بسیار پیچیده می‌باشد؛ بنابراین استفاده از یک مدل کلاسیک به‌تنهایی نمی‌تواند پیش‌بینی دقیقی از شاخص‌های قیمت سهام داشته باشد. ازاین‌رو با به‌کارگیری روش‌های ترکیبی می‌توان عدم اطمینان در پیش‌بینی را کاهش داد (دولو و صفری، ۱۳۹۵).

۲- مبانی نظری و پیشینه پژوهش

قیمت سهام یک منبع اطلاعاتی مهم و مؤثر جهت تصمیم‌گیری برای افراد حقیقی و حقوقی در بازار سهام محسوب می‌شود. به‌تبع آن پیش‌بینی تغییرات قیمت سهام نیز از اهمیت بالایی برخوردار خواهد بود. با وجود ابزارهای تحلیلی و ریاضی جهت پیش‌بینی قیمت سهام عامل اول تعیین شاخص‌های مؤثر بر آن است. در این راستا مطالعات متعددی صورت‌گرفته است و اجماعی روی آنها وجود ندارد. در این بخش از مقاله به تشریح برخی از پژوهش‌های انجام شده در زمینه پیش‌بینی قیمت سهام خواهیم پرداخت.

کیموتو و همکاران[1] (۱۹۹۰) در پژوهشی تحت عنوان "سیستم پیش‌بینی بازار سهام با شبکه‌های عصبی مدولار"، به یک سیستم پیش‌بینی پرداخته‌اند که زمان‌بندی خریدوفروش سهام را توصیه می‌کند. در نهایت به این نتیجه رسیدند که سیستم پیش‌بینی فعلی از بازده‌های آینده برای تولید داده‌های آموزشی استفاده می‌کند. سیستمی که در آن داده‌های تدریس به همراه یک روش آماری تولید می‌شود، باید توسعه یابد.

1. Kimoto *et al.*

جنا و پادهی (۲۰۱۴) در پژوهشی تحت عنوان "کاربرد GA با SVM برای پیش‌بینی قیمت سهام در بازار مالی" از الگوریتم GA-SVM برای پیش‌بینی شاخص قیمت سهام استفاده می‌کند. نتایج تجربی نشان داد که SVM با GA جایگزین بهینه و امیدوار کننده‌تری برای پیش‌بینی بازار سهام فراهم می‌کند.

فلاح پور و همکاران (۱۳۹۲) در پژوهشی تحت عنوان "پیش‌بینی روند حرکتی قیمت سهام با استفاده از ماشین بردار پشتیبان برپایهٔ الگوریتم ژنتیک در بورس اوراق بهادار تهران" برای نمونهٔ آماری، سی شرکت از پنجاه شرکت برتر بورس اوراق بهادار در سه ماههٔ دوم سال ۹۰ انتخاب کردند. سپس برای هر سی شرکت، ۴۴ متغیر محاسبه شد. نتایج نشان داد، 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.

سروانتس و همکاران[1] (۲۰۱۵) در پژوهشی تحت عنوان "انتخاب داده‌ها بر اساس درخت تصمیم برای طبقه‌بندی SVM در مجموعه داده‌های بزرگ" الگوریتم جدیدی برای سرعت بخشیدن به زمان آموزش SVM ارائه داده‌اند. نتایج نشان داد که الگوریتم پیشنهادی نتایج با دقت مشابه و به روش سریعتر از اجرای SVM فعلی تولید می‌کند. عبادتی و مرتضوی (۲۰۱۸) در پژوهشی تحت عنوان "یک روش کارا مبتنی بر یادگیری ماشین برای پیش‌بینی سری‌های زمانی در بازار سهام" از روش ترکیبی الگوریتم ژنتیک (GA) و تکنیک شبکه عصبی مصنوعی (ANN[2]) برای توسعه روشی به‌منظور پیش‌بینی قیمت سهام وسری‌های زمانی استفاده می‌کند. نتایج به‌دست‌آمده در پژوهش نشان داد که روش پیشنهادی پژوهش دارای دقت بسیار بالاتری نسبت به‌دقت هرکدام از روش‌های قبل می‌باشد.

ونگ و همکاران (۲۰۱۷) در پژوهشی تحت عنوان "پیش‌بینی حرکت یک‌روزه بازار سهام با استفاده از اختلاف منابع داده" فرض کردند که ترکیب منابع داده‌های آنلاین متفاوت با شاخص‌های سری زمانی سنتی و شاخص‌های فنی برای سهام می‌تواند یک سیستم خبره تجارت روزانه مؤثرتر و باهوش‌تر را فراهم کند. سرانجام، یک ابزار تصمیم‌گیری هوشمند برای کمک به سرمایه‌گذاران در تصمیم‌گیری‌های تجاری در مورد هر بورس، کالا یا شاخص ارائه دادند.

اقبال و همکاران[3] (۲۰۱۳) در پژوهشی تحت عنوان "تکنیک‌های یادگیری ماشین کارآمد برای پیش‌بینی بازار سهام" از روش مقایسه‌ای برای یافتن تکنیک بهینه برای پیش‌بینی بازار سهام استفاده

1. Cervantes *et al.*
2. Artificial Neural Network
3. Iqbal *et al.*

کردند و دریافتند که شبکه عصبی مکرر (RNN[1]) عملکرد بیشتری نسبت به شبکه عصبی مصنوعی (ANN) برای پیش‌بینی دارد و شبکه عصبی مکرر لایه‌ای (LRNN) عملکرد بهتری نسبت به شبکه عصبی FeedForward (NN) دارد.

ژانگ و همکاران[2] (۲۰۱۹) در پژوهشی تحت عنوان "پیش‌بینی سهام بورس بر اساس شبکه مخالف تولیدی" معماری anovel شبکه (GAN [3]) را با چندلایه Perceptron (MLP) به‌عنوان تمایز دهنده و حافظه کوتاه‌مدت – بلندمدت (LSTM[4]) به‌عنوان تولیدکننده پیش‌بینی قیمت بسته‌شدن سهام پیشنهاد می‌کند. نتایج تجربی نشان می‌دهد که رمان GAN می‌تواند عملکردی امیدوارکننده در پیش‌بینی قیمت بسته‌شدن داده‌های واقعی در مقایسه با سایر مدل‌های یادگیری و یادگیری عمیق داشته باشد.

حسین‌زاده و هراتی‌زاده[5] (۲۰۱۹) در پژوهشی تحت عنوان "پیش‌بینی بازار سهام مبتنی بر CNN[6] با استفاده از مجموعه متنوعی از متغیرها" یک چارچوب مبتنی بر CNN را پیشنهاد می‌کند. ارزیابی‌ها نشان دهنده بهبود قابل توجهی در عملکرد پیش‌بینی در مقایسه با وضعیت پایه‌های اولیه هنر است.

۳- روش‌شناسی پژوهش

فرضیه‌های پژوهش به‌صورت زیر می‌باشند که در ادامه به اثبات آن‌ها می‌پردازیم:

الگوریتم ترکیبی GA-SVM عملکرد بهتری نسبت به الگوریتم SVM دارد.

شاخص‌های NMSE, MAE, DS شاخص‌های مناسبی برای پیش‌بینی قیمت سهام می‌باشند.

الگوریتم SVM یک ابزار محاسبات ریاضی می‌باشد که مبتنی بر اصل حداقل‌سازی عملیاتی است و سابقه آن به سال ۱۹۶۰ برمی‌گردد. الگوریتم ماشین بردار SVM بر اساس نظریه یادگیری آماری بنا شده و یک روش آماری غیر پارامتریک نظارت شده است (لامبا و کومار[7]، ۲۰۱۶). ماشین بردار پشتیبان SVM به‌عنوان یکی از قدیمی‌ترین و دقیق‌ترین روش‌های یادگیری ماشینی در میان الگوریتم‌های معروف شناخته می‌شود. الگوریتم SVM ازالگوریتم‌های تشخیص الگوی دسته‌بندی می‌باشد. از این الگوریتم در تشخیص الگو یا دسته‌بندی اشیاء در کلاس‌های خاص می‌توان استفاده کرد. همچنین ماشین بردار پشتیبان SVM یکی از روش‌های یادگیری با ناظر

1. Recurrent Neural Network
2. Zhang *et al*.
3. Generator Adversarial Network
4. Long Short-Term Memory
5. Hoseinzade & Haratizadeh
6. Cable News Network
7. Lamba & Kumar

است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. این روش تا حدودی پیچیده است و ویژگی مثبت آن در این است که به تعداد نمونه‌های آموزش وابسته نمی‌باشد و با تعداد ویژگی‌های بالا و تعداد نمونه‌های کم می‌تواند به خوبی کار کند (فهمی و همکاران، ۱۳۹۷).

مدل رگرسیونی SVM تابعی است که مرتبط با متغیر وابسته Y می‌باشد و خود تابعی از چند متغیر مستقل X است. مانند سایر مسئله‌های رگرسیونی، در اینجا فرض بر این است که تابع جبری $f(x)$ رابطه میان متغیرهای مستقل و وابسته باشد که با ضابطه $(f(x) = w^T\ \phi(x) + b)$ به علاوه مقداری اغتشاش[1] مشخص شود $(y = f(x) + noise)$. شایان ذکر است که در مراجع مربوط با الگوریتم SVM این مقدار اغتشاش به‌عنوان خطای مجاز (ε) تعریف شده است. چنانچه w (بردار ضریب) و b (ثابت) مشخصه‌های تابع رگرسیونی و ϕ نیز تابع کرنل باشد، آنگاه هدف پیدا کردن فرم تابعی برای $f(x)$ است. این مهم با آموزش مدل SVM توسط مجموعه ای از نمونه‌ها (مجموعه آموزش) محقق می‌شود. این روند شامل بهینه‌سازی متوالی تابع خطاست. بسته به تعریف این تابع خطا دو مدل SVM تعریف می‌شود:

SVM رگرسیونی نوع ۱ (به‌عنوان $\varepsilon - SVM$ رگرسیونی نیز شناخته می‌شود).

SVM رگرسیونی نوع ۲ (به‌عنوان $\upsilon - SVM$ رگرسیونی نیز شناخته می‌شود).

شایان‌ذکر است که در این تحقیق از مدل $\varepsilon - SVM$ رگرسیونی، به دلیل کاربرد گسترده آن در مطالعات رگرسیونی، برای تخمین به هنگام BOD_5 استفاده شده است؛ بنابراین برای محاسبه w و b لازم است تابع خطا (رابطه ۱) در مدل $\varepsilon - SVM$ با در نظر گرفتن شرایط مندرج در رابطه ۲ بهینه شود.

رابطه (۱)
$$\frac{1}{2}w^T.w + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\xi_i^*$$

$$w^T.\phi(x_i) + \mathrm{b} - y_i \le \varepsilon + \xi_i^*$$

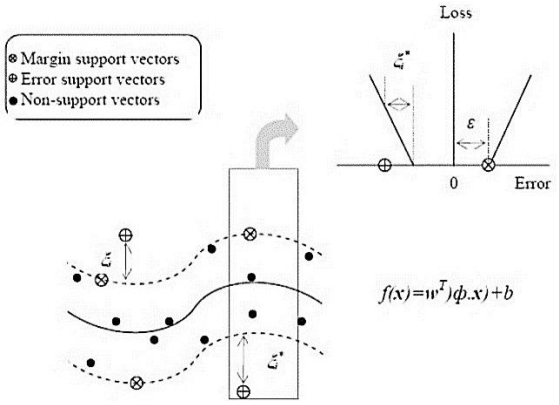
$$y_i - w^T.\phi(x_i) - \mathrm{b} \le \varepsilon + \xi_i$$

رابطه (۲)
$$\frac{1}{2}w^T.w + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\xi_i^*$$

C عددی صحیح مثبت می‌باشد که در معادلات بالا در هنگام رخدادن خطای آموزش مدل، عامل تعیین جریمه است. متغیرهای کمبود[2] شامل تابع کرنل[3] ϕ ، N تعداد نمونه‌های مدل و دو مشخصه ξ_i

1. Noise
2. Deficiency Variables
3. Kernel Function

و ξ_i^* هستند که حدود بالا و پایین خطای آموزش مرتبط با مقدار خطای مجاز ε را مشخص می‌کنند. پیش‌بینی می‌شود که در مسائل، داده‌ها درون بازه مرزی ε قرار گیرند (شکل ۱). اگر داده‌ای خارج از بازه ε قرار گرفت، آنگاه یک خطا معادل ξ_i و ξ_i^* وجود خواهد داشت. ذکر این نکته نیز لازم است که مدل SVM مشکلات ناشی از کم تخمینی[1] و فوق برازشی[2] را با کمینه کردن همزمان دو ترم $w^T.w/2$ و خطای آموزشی، یعنی $C\sum_{i=1}^{N}(\xi_i+\xi_i^*)$ را در رابطه ۱ حل می‌کند؛ بنابراین مسئله بهینه‌سازی با حداکثرسازی عددی تابع درجه دوم زیر (رابطه ۳) به‌وسیله ۲ ضریب لاگرانژ a_i و a_i^* با شرایط رابطه ۴ حل خواهد شد.



شکل (۱) مدل رگرسیونی SVM (اسکندری و همکاران، ۱۳۹۱)

Figure (1) SVM regression model (Eskandari *et al*, 2012)

رابطه (۳)
$$\sum_{i=1}^{N} y_i\,(a_i-a_i^*)-\varepsilon\sum_{i=1}^{N}(a_i-a_i^*)-0.5\sum_{i,j=1}^{N}(a_i-a_i^*)\left(a_j-a_j^*\right)\phi(x_i)^T\,\phi(x_j)$$

رابطه (۴)
$$\sum_{i=1}^{N}(a_i-a_i^*)=0$$

$$0\le a_i\le C$$

$$0\le a_i^*\le C \quad , \quad i=1,\dots,N$$

جواب رابطه ۳ یکتا خواهد بود چون تابع هدف بالا در رابطه ۳ تابع محدب است. پس از اینکه ضرایب لاگرانژ در رابطه ۳ را تعریف کردیم، متغیرهای w و b در مدل رگرسیونی SVM به‌وسیله

1. Slack Variable
2. Over Fitting

نظریه کراش-کوهن-تاکر[1] محاسبه می‌شوند (فلچر[2]، ۱۹۸۷). که در آن $W = \sum_{i=1}^{N}(a_i - a_i^*)\phi(x_i)$ است. در نتیجه برای مدل SVM رگرسیونی خواهیم داشت:

$$W = \sum_{i,=1}^{N}(a_i - a_i^*)\phi(x_i)^T\,\phi(x) + b \qquad \text{رابطه (۵)}$$

باید توجه داشت که ترم‌های لاگرانژ $(a_i - a_i^*)$ می تواند صفر و یا غیر صفر باشند؛ بنابراین فقط مجموعه داده‌هایی در معادله رگرسیون نهایی وارد می‌شوند که ضرایب a_i^- آنها غیر صفر است. این مجموعه داده ها، بردارهای پشتیبان می‌باشند. به بیان دیگر، بردارهای پشتیبان داده‌هایی هستند، که باعث به وجود آمدن تابع رگرسیونی می‌شوند. بردارهای پشتیبان حاشیه‌ای[3] ، آن بردارهایی هستند که مقدار $|a_i^-|$ آنها کمتر از C باشد. بردار پشتیبان خطا[4]، یا بردار پشتیبان کراندار بردارهایی هستند که مقدار $|a_i^-|$ آنها برابر مقدار C باشد. بردارهای پشتیبان حاشیه‌ای در حاشیه مرز غیرحساس یافت می‌شوند. در حالی که بردارهای پشتیبان خطا خارج از بازه هستند (شکل ۳). در نهایت تابع SVM رگرسیونی را می‌توان به فرم زیر بازنویسی کرد:

$$f(x) = \sum_{i,=1}^{N} a_i^-.\phi(x_i)^T\,\phi(x) + b \qquad \text{رابطه (۶)}$$

در رابطه ۶ محاسبه $\phi(x)$ در فضای مشخصه آن ممکن است بسیار پیچیده باشد. برای حل این مشکل روند معمول در مدل SVM رگرسیونی انتخاب یک تابع کرنل به صورت $K(x_i, x) = \phi(x_i)^T\sqrt{b^2 - 4ac}$ است. با استفاده از عبارت w و b ، $f(x)$ به صورت زیر محاسبه می‌شود:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\ \Phi(x_i).\Phi(x) + b = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\,k(x_i.x) + b \qquad \text{رابطه (۷)}$$

لازم به ذکر است که ما به تابع Φ نیاز نداریم که $f(x)$ را محاسبه کنیم که یکی از مزایای استفاده از کرنل است. می‌توان از توابع مختلف کرنل برای ساخت انواع مختلف $\varepsilon - SVM$ استفاده کرد. انواع رایج توابع کرنل قابل استفاده در مدل SVM رگرسیونی عبارتند از: کرنل چندجمله ای با سه مشخصه هدف، کرنل سیگموییدی[5] شامل دو مشخصه هدف و کرنل توابع پایه شعاعی[6] (RBF) با یک مشخصه هدف (اسکندری و همکاران، ۱۳۹۱).

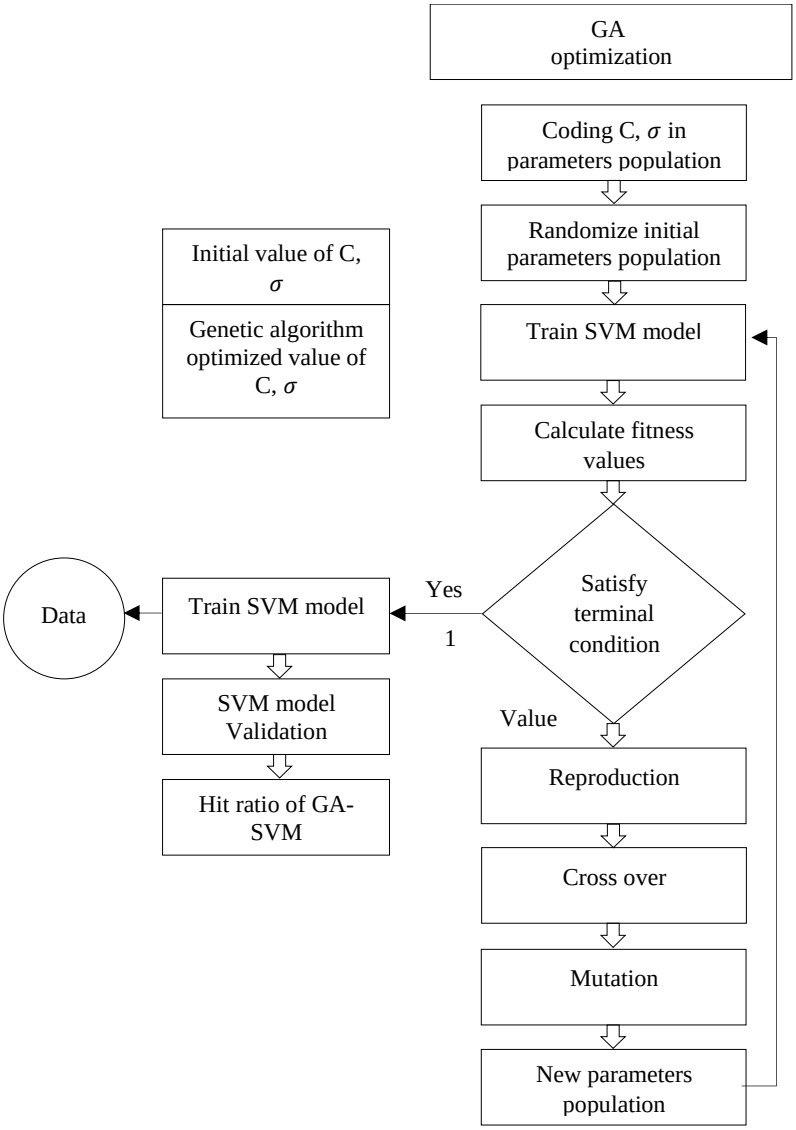
1. Karush – Kuhn - Tucker
2. Fletcher
3. Margin Support Vector
4. Error Support Vector
5. Sigmoid Kernel
6. Radial Basis Function (RBF)

الگوریتم ژنتیک، برای اولین‌بار توسط هلند (۱۹۷۵) استفاده و به‌عنوان ابزار قوی بهینه‌سازی توسعه پیدا کرد. GA یک الگوریتم جستجو، برگرفته از طبیعت بیولوژیکی و فرآیند انتخاب طبیعی می‌باشد. اساس این روش مبتنی بر نظریه داروین که در محیط متغیر همواره موجوداتی ادامه حیات دارند که از همه پایدارترند، می‌باشد (البرزی، ۱۳۹۳).

الگوریتم GA به‌وسیله مجموعه‌ای از جواب‌های تصادفی اولیه (جمعیت) شروع می‌شود. هر جمعیت از مجموعه‌ای از کروموزوم‌ها که هرکدام یک جواب مسئله هستند، تشکیل می‌شود و هر کروموزوم مجموعه‌ای از ژن‌ها است که همان متغیرهای تصمیم مسئله می‌باشند. عملکرد الگوریتم GA از اندازه جمعیت تأثیر می‌پذیرد. اگر تعداد جمعیت خیلی کم باشد، به‌خاطر عدم جستجو در تمام فضای جواب، ممکن است الگوریتم به جواب مطلوب همگرا نگردد و اگر تعداد جمعیت زیاد باشد، سرعت همگرایی به سمت جواب بهینه کند خواهد بود، زیرا فضای بیشتری جستجو می‌شود.

دو نوع عملگر در الگوریتم GA موجود است، عملگرهای تکاملی همان عملگر انتخاب و عملگرهای ژنتیک همان عملگر جابه‌جایی و جهش می‌باشند. فرآیند انتخاب کروموزوم بر اساس میزان شایستگی توابع هدف متناظر با هر کروموزوم در هر نسل می‌باشد و معیار انتخاب کروموزوم‌ها بر اساس شایستگی آن‌ها است. می‌توان از دو روش برای ایجاد نسل بعدی که همان فرزندان می‌باشند، استفاده کرد: ترکیب دو کروموزوم، به‌وسیله عملگر جابه‌جایی و اصلاح برخی کروموزوم‌ها به‌وسیله عملگر جهش ژنی، در نهایت تولید نسل تا جایی ادامه می‌یابد که معیار توقف حاصل شود و مقدار تابع هدف کمینه شود (آذرافزا و همکاران، ۱۳۹۱).

در این پژوهش، ما بهترین ورودی‌ها را برای پیش‌بینی مدل SVM از طریق الگوریتم GA از مجموعه‌داده‌های ورودی انتخاب می‌کنیم. الگوریتم ژنتیک بکار رفته در این پژوهش یک روش بهینه‌سازی مبتنی بر جمعیت، بهینه‌سازی تکاملی را ارائه می‌دهد که به‌موجب آن جمعیت‌های تعریف شده با استفاده از اصل داروینی بقا از طریق نسل‌ها تکامل می‌یابند. اگر ویژگی‌های لازم به‌شدت وجود داشته باشد. GA بهترین ویژگی را برای حل مشکلات بهینه‌سازی ارائه می‌دهد، در شکل ۲ نیز طرز کار الگوریتم GA نشان‌داده‌شده است، برای انتخاب C و δ ابتدا اعداد تصادفی را انتخاب می‌کند و روی آن‌ها الگوریتم را پیاده‌سازی می‌کند و به الگوریتم SVM ارسال می‌کند درصورتی‌که جواب مناسب نباشد دوباره روی آن‌ها کار می‌کند و این انتخاب‌ها تا جایی ادامه پیدا می‌کند که جواب بهینه به دست آید و بهترین دقت پیش‌بینی به دست آید.



شکل (۲) فلوچارت انتخاب پارامترهای بهینه توسط الگوریتم GA (جنا و پادهی، ۲۰۱۴)

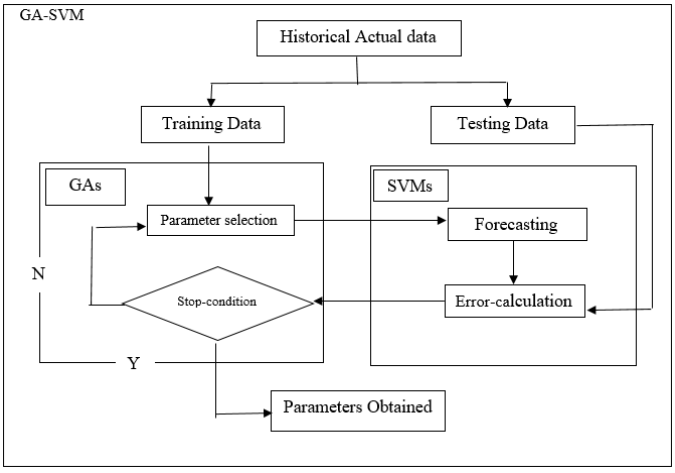
Figure (2) Flowchart of selection of optimal parameters by genetic algorithm (Jena & Padhy, 2014)

باتوجه‌به اینکه که ماهیت بازارها در مناطق مختلف متفاوت است، ما از دو روش یادگیری ماشین: پشتیبانی از تکنیک ماشین بردار (SVM) و تکنیک ماشین بردار پشتیبانی با الگوریتم ژنتیکی (GA-SVM) برای پیش‌بینی قیمت‌های آتی معامله شده سهام ۸ شرکت بین‌المللی استفاده کردیم. در این

پژوهش کار بهینه‌سازی، با استفاده از الگوریتم GA، برای تحقق و ازبین‌بردن ورودی‌هایی که برای پیش‌بینی لازم نیست و یا اجتناب از آسیب‌پذیری مناسب است که معمولاً در مدل‌هایی با پارامترهای خیلی زیاد اتفاق می‌افتد، حاصل می‌شود. این امر منجر به بهبود سرعت پردازش SVM شده و دقت پیش‌بینی را افزایش می‌دهد.

مدل ترکیبی ارائه‌شده در این پژوهش شامل دو مرحله می‌باشد. در مرحله اول با استفاده از الگوریتم GA جمعیت بهینه از داده‌های روزانه را انتخاب می‌کنیم و پس از آن در مرحله دوم داده‌های بهینه شده به‌وسیله الگوریتم GA را به الگوریتم SVM می‌دهیم تا به‌وسیله آن‌ها به پیش‌بینی شاخص قیمت سهام بپردازد.

در این قسمت نحوه کار الگوریتم SVM و GA-SVM توضیح داده خواهد شد. در شکل ۳ معماری مدل GA-AVM آورده شده است.



شکل (۳) معماری مدل GA-SVM (جنا و پادهی، ۲۰۱۴)

Figure (3) A-SVM model architecture (Jena & Padhy, 2014)

در مدل ترکیبی GA-SVM با ترکیب نتایج هر دو مدل به پیش‌بینی قیمت سهام می‌پردازیم، الگوریتم SVM برای پیش‌بینی نیاز به دو مقدار عددی اولیه C و δ می‌باشد. در پژوهش‌های قبلی که به‌وسیله الگوریتم SVM به پیش‌بینی پرداخته شده بود، این اعداد را به‌صورت تصادفی به الگوریتم SVM می‌دادند و به پیش‌بینی قیمت می‌پرداختند. اما در مدل ترکیبی GA-SVM مقادیر C و δ به‌وسیله الگوریتم GA به دست می‌آید و به الگوریتم SVM داده می‌شود تا عمل پیش‌بینی قیمت انجام شود. این عمل آن‌قدر تکرار می‌شود تا به جواب موردنظر برسیم و جواب بهینه را به دست آوریم. پیش‌ازاین پیش‌بینی سری زمانی توسط کیونگ جه کیم از کره جنوبی با درنظرگرفتن SVM انجام

شده بود. در SVM برخی از مقادیر خاص دو پارامتر C و δ انتخاب شدند. برای یافتن پیش‌بینی بهتر، از مقادیر C و δ بین مقادیر به‌صورت اعداد احتمالی استفاده کرد. در اینجا از تکنیک ترکیبی GA-SVM به‌منظور بهبود دقت پیش‌بینی با انجام انتخاب بهینه پارامتر C و δ استفاده شده است.

برای راه‌اندازی آزمایش و برای اثبات صحت پیش‌بینی روش پیشنهادی ما، داده‌های سری زمانی برخی از شرکت‌های بین‌المللی را گرفته‌ایم و پیش‌بینی قیمت سهم آن‌ها را با استفاده از SVM محاسبه کرده‌ایم. باز هم پارامترهایی که در SVM استفاده می‌شوند، در الگوریتم GA پیشنهادی منتقل می‌شوند تا مقادیر بهینه حاصل شود. همان داده‌ها نیز برای مقایسه دقت پیش‌بینی قیمت سهم با استفاده از الگوریتم GA-SVM و SVM پیاده‌سازی شده است.

برای تجزیه‌وتحلیل داده‌ها می‌توان از سری‌های زمانی به‌دست‌آمده از بازارهای مالی و تکنیک‌های داده‌کاوی استفاده کرد و به پیش‌بینی قیمت سهام پرداخت. پیش‌بینی سری زمانی مستلزم به‌دست‌آوردن شاخص‌هایی از داده‌های سری زمانی است. بیش از ۱۰۰ نوع شاخص برای پیش‌بینی قیمت سهام در بازارهای مالی توسعه داده شده است تا بتوان رفتار بازار را طریق آن‌ها تشخیص داد. تعیین شاخص‌های مناسب از نکات مهم مسائل پیش‌بینی سری زمانی سهام است.

در پژوهش حاضر عملکرد پیش‌بینی سهام را با استفاده از معیارهای آماری جدول ۱، یعنی میانگین خطای میانگین مربعات نرمال (NMSE)، میانگین خطای مطلق (MAE) و تقارن جهت (DS) ارزیابی می‌کنیم. تعاریف این معیارها در جدول ۱ یافت می‌شود. NMSE و MAE میزان انحراف بین قیمت واقعی و پیش‌بینی شده هستند. هر چه مقادیر NMSE و MAE کوچک‌تر باشند، قیمت‌های پیش‌بینی شده به قیمت‌های واقعی نزدیک‌تر هستند. DS نشانه‌ای از صحت جهت پیش‌بینی شده قیمت‌ها می‌باشد. (مقدار زیادی بیانگر پیش‌بینی‌کننده بهتر می‌باشد.)

جدول (۱) محاسبه شاخص‌های سری زمانی

Table (1) Calculation of time series indices

شاخص	محاسبه
NMSE	$\frac{1}{\sigma^2 n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ $\sigma^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2\ \hat{y}=\sum_{i=1}^{n}y_i$
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert$
DS	$\frac{100}{n}\sum_{i=1}^{n}d_i$ $d_i=\begin{cases}1=(y_i-y_{i-1})-(\hat{y}_i-\hat{y}_{i-1}) & \geq 0 \\ 0 & otherwise\end{cases}$

عملکرد پیش‌بینی با استفاده از پارامترهای NMSE(Namely ,the normalized Mean Squared Error)، MAE(Mean Absolute Error) و DS(Directional Symmetry) ارزیابی می‌شوند. N، تعداد کل الگوهای داده؛ y، مقدار واقعی؛ $\hat{y}$، مقدار پیش‌بینی شده، NMSE و MAE انحراف معیار مقادیر واقعی و پیش‌بینی هستند. کوچکترین مقادیر NMSE وMAE مقدار پیش‌بینی شده نزدیک‌تری به مقدار واقعی هستند. DS به علامت درستی جهت پیش‌بینی به صورت درصدی اشاره دارد. (مقدار بیشتر پیشگویی بهتری را پیشنهاد می‌دهد.)

در این قسمت نیز شبه کد مربوط به الگوریتم‌های استفاده شده که در این پژوهش، از طریق نرم‌افزار متلب اجرا و پیاده سازی شده‌اند، را نشان خواهیم داد که به صورت زیر می‌باشند، شبه کدهای الگوریتم SVM و GA-SVM را به صورت جداگانه در دو شکل ۴ و شکل ۵ نشان داده شده است.

```
%
%SVM
model_SVM = fitrsvm(train(:,1:end1),train(:,end),'Standardize',true,'KernelFunction',kernel);
ySVM = predict(model_SVM,test(:,1:end-1));
```

شکل (۴) شبه کد الگوریتم SVM

Figure (4) Pseudo-code of SVM algorithm

```
%%
%SVMG
parameter =
GeneticAlgorithm(train,test,PopSise,crossover_rate,mutation_rate,MaxIter,mode,kernel);
model_SVMG = fitrsvm(train(:,1:end-1),train(:,end),...
'Standardize',true,'KernelFunction',kernel,...
'BoxConstraint',parameter(1),'KernelScale',parameter(2),'Epsilon',parameter(3));
Ysvmg = predict(model_SVMG,test(:,1:end);
```

شکل (۵) شبه کد الگوریتم ترکیبی GA-SVM

Figure (5) Pseudo-code of the Hybrid GA-SVM algorithm

٤- تجزیه‌وتحلیل داده‌ها

مدل پیش‌بینی با داده‌های آتی شاخص واقعی زیر که از بورس جمع‌آوری شده، استفاده می‌شود. آن‌ها داده‌های ۸ شرکت بین‌المللی که شامل شرکت Apple, Google, Facebook, IBM, Ford Yahoo, Microsoft, Twitter, هستند. ما برای هر یک از قراردادهای آتی که در بالا ذکر شد ۱۸۰ نمونه گرفته‌ایم. هر قرارداد از ۱ ژانویه ۲۰۱۶ شروع شده است. داده‌های جمع‌آوری شده شامل قیمت

روزانه بسته‌شدن سایت، قیمت باز، قیمت بالا، قیمت پایین، حجم معامله شده و ارزش معاملات است. قیمت بسته‌شدن روزانه به‌عنوان مجموعه‌داده‌ها استفاده می‌شود.

نتایج پیش‌بینی الگوریتم SVM و الگوریتم ترکیبی GA-SVM برای مجموعه آزمون در جدول ۲ جمع‌آوری شده است که نشان می‌دهد الگوریتم ترکیبیGA-SVM در اکثر موارد بهتر از الگوریتم SVM می‌باشد. الگوریتم ترکیبی GA-SVM در بیشتر موارد NMSE و MAE کمتر و DS بزرگ‌تر را نسبت به الگوریتم SVM فراهم می‌کند. در این پژوهش برای سنجیدن دقت الگوریتم‌های SVM و GA-SVM از نتایج شاخص‌های NMSE و MEA و DS استفاده می‌کنیم، NMSE و MEA هرچه مقدارشان کمتر باشد، پیش‌بینی نتیجه مناسب‌تری می‌دهد و همچنین شاخص DS نیز هرچه بزرگ‌تر باشد نتیجه پیش‌بینی بهتر می‌باشد. پس از اجرای هر دو مدل SVM و GA-SVM و باتوجه‌به نتایجی که در نمودارها به‌دست‌آمده و همچنین مقادیری که در جدول ۲ آمده، نشان می‌دهد که دقت الگوریتم ترکیبی GA-SVM خیلی بهتر از دقت الگوریتم SVM می‌باشد. شاخص‌های عملکرد تعیین شده برای آزمایش ما وقتی از روش ترکیبی GA-SVM در مقایسه با SVM استفاده می‌شود، توافق بسیار خوبی از قیمت پیش‌بینی شده با قیمت واقعی نشان دادند.

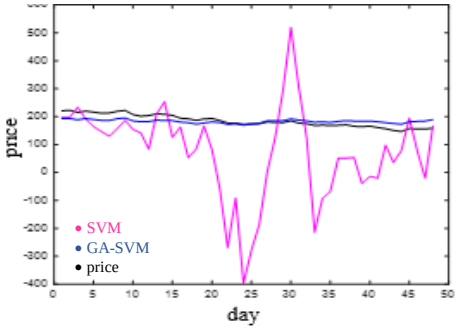
جدول (۲) نتایج عددی شاخص‌های سری زمانی توسط الگوریتم‌های SVM و GA-SVM برای ۱۸۰ روز

Table (2) Numerical results of time series indices by SVM and Hybrid GA-SVM algorithms for 180 days

شاخص	NSME		MAE		DS	
	GA-SVM	SVM	GA-SVM	SVM	GA-SVM	SVM
Apple	۷۴۸۱/۰	۵۸۳۷/۷۳	۱۴۸۶/۱۷	۶۳۱۶/۱۳۹	۵/۸۷	۵/۶۲
Yahoo	۳۸۹۹/۵	۵۹۸۴/۷۸	۸۸۲۲/۲	۸۲۵۶/۱۰	۰/۷۵	۰/۵۰
Google	۸۰۵۳/۰	۳۷۹۶/۴	۷۴۱۱/۱۳	۳۵۳۵/۲۹	۸۳۳۳/۷۰	۷۵۰۰/۴۳
Facebook	۱۶۱۸/۲	۷۵۹۸/۷۹	۶۶۱۱/۷	۶۷۶۲/۳۱	۹۱۶۷/۷۲	۲۵۰۰/۵۶
Ibm	۵۰۷۳/۱	۳۱۸۷/۴	۹۰۶۲/۲	۴۲۵۹/۴	۹۱۶۷/۷۲	۶۶۶۷/۶۶
Microsoft	۸۲۹۸/۰	۳۴۶۹/۳	۰۲۷۸/۱	۹۰۱۸/۱	۸۳۳۳/۷۰	۷۵۰۰/۴۳
Twitter	۳۳۶۵/۰	۱۵۲۶/۱۱	۸۰۹۲/۰	۳۰۲۹/۴	۳۳۳۳/۸۳	۴۱۶۷/۶۰
Ford	۸۰۰۸/۰	۶۳۶۷/۱۴	۵۲۱۴/۰	۸۹۶۴/۱	۲۵/۸۱	۶۶۶۷/۶۶

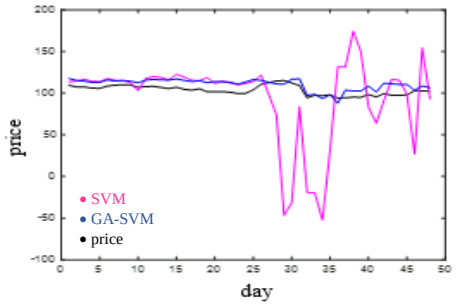
پس از آموزش با رگرسیون ماشین بردار پشتیبان SVM و ماشین بردار پشتیبان با رگرسیون GA، قیمت پیش‌بینی‌شده و قیمت واقعی برای داده‌های تست در شکل‌های ۱۳-۶ به نمایش گذاشته می‌شوند. همان‌طور که در شکل‌های ۱۳-۶ مشاهده می‌کنیم دقت پیش‌بینی برای مدت ۱۸۰ روز، دقت مناسب و قابل‌قبول می‌باشد. نمودار الگوریتم SVM نوسانات زیادی دارد، اما نمودار مربوط به الگوریتم ترکیبی GA-SVM به داده‌های واقعی نزدیک‌تر می‌باشد و دقت بسیار بالاتری دارد. یکسری از داده‌ها

برای آموزش، مابقی برای تست می‌باشند. در قسمت آموزش داده‌ها، نمودارها حرکت تقریباً مشابهی دارند. اما در قسمت تست نمودار، الگوریتم ترکیبی GA-SVM پیش‌بینی بسیار بهتری نسبت به الگوریتم SVM را دارد.



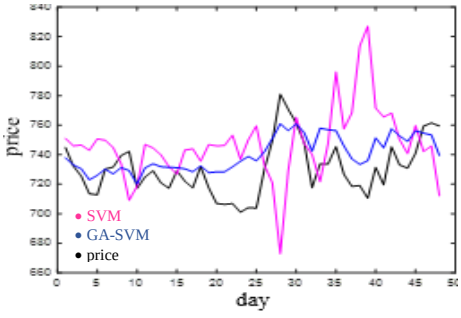
شکل (۷) نتایج پیش‌بینی توسط الگوریتم SVM و الگوریتم ترکیبی GA-SVM شرکت Facebook

Figure (7) Prediction results by SVM algorithm and Hybrid GA-SVM algorithm from Facebook Company



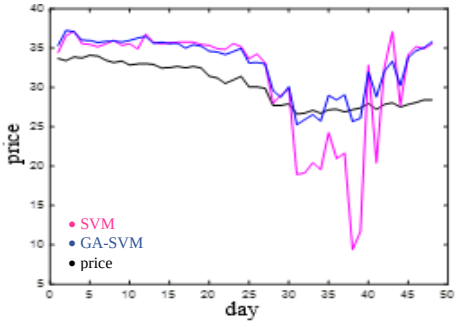
شکل (۶) نتایج پیش‌بینی توسط الگوریتم SVM و الگوریتم ترکیبی GA-SVM شرکت Apple

Figure (6) Prediction results by SVM algorithm and Hybrid GA-SVM algorithmm from Apple Company



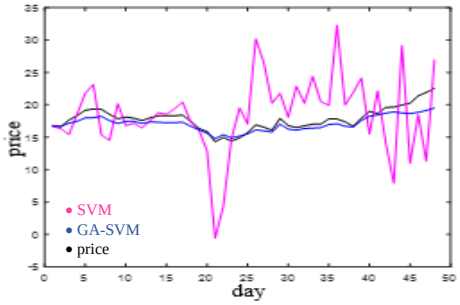
شکل (۹) نتایج پیش‌بینی توسط الگوریتم SVM و الگوریتم ترکیبی GA-SVM شرکت IBM

Figure (9) Prediction results by SVM algorithm and Hybrid GA-SVM algorithm from IBM Company



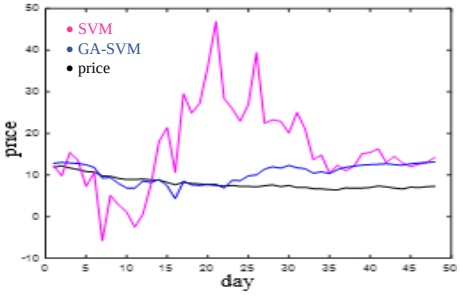
شکل (۸) نتایج پیش‌بینی توسط الگوریتم SVM و الگوریتم ترکیبی GA-SVM شرکت Google

Figure (8) Prediction results by SVM algorithm and Hybrid GA-SVM algorithm from Google Company



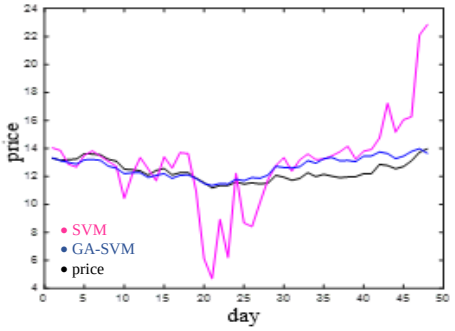
شکل (۱۰) نتایج پیش‌بینی توسط الگوریتم SVM و الگوریتم ترکیبی GA-SVM شرکت Yahoo

Figure (10) Prediction results by SVM algorithm and Hybrid GA-SVM algorithm from Yahoo Company



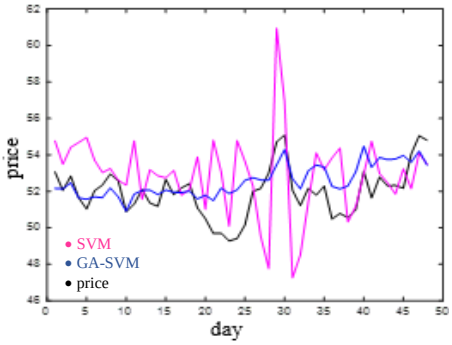
شکل (۱۱) نتایج پیش‌بینی توسط الگوریتم SVM و الگوریتم ترکیبی GA-SVM شرکت Twitter

Figure (11) Prediction results by SVM algorithm and Hybrid GA-SVM algorithm from Twitter Company



شکل (۱۲) نتایج پیش‌بینی توسط الگوریتم SVM و الگوریتم ترکیبی GA-SVM شرکت Microsoft

Figure (12) Prediction results by SVM algorithm and Hybrid GA-SVM algorithm from Microsoft Company



شکل (۱۳) نتایج پیش‌بینی توسط الگوریتم SVM و الگوریتم ترکیبی GA-SVM شرکت Ford

Figure (13) Prediction results by SVM algorithm and Hybrid GA-SVM algorithm from Ford Company

۵- بحث و نتیجه‌گیری

باتوجه‌به مقادیر شاخص‌ها در جدول ۲ و همچنین شکل‌های ۱۳-۶ الگوریتم SVM و GA-SVM برای همه شرکت‌هایی که در اینجا بررسی کردیم می‌توان نتیجه گرفت که دقت پیش‌بینی الگوریتم ترکیبی GA-SVM بهتر از دقت الگوریتم SVM می‌باشد و به قیمت واقعی نزدیک‌تر می‌باشد. در تمامی نمودارهای شرکت‌ها، حرکت قیمت الگوریتم‌های SVM و GA-SVM در قسمت آموزش داده‌ها، تقریباً یکسان می‌باشد، اما در قسمت تست داده‌ها، نمودار الگوریتم ترکیبی GA-SVM بهتر از الگوریتم SVM می‌باشد و به قیمت واقعی نزدیک‌تر می‌باشد.

در پژوهش حاضر، با استفاده از الگوریتم ماشین بردار پشتیبان SVM و الگوریتم ماشین بردار پشتیبان با رگرسیون GA که همان الگوریتم ترکیبی GA-SVM نام دارد، به پیش‌بینی قیمت سهام ۸ شرکت بین‌المللی در طی ۱۸۰ روز پرداختیم و مشاهده نمودیم که الگوریتم ترکیبی GA-SVM دقت پیش‌بینی بسیار بالاتری نسبت به الگوریتم SVM دارد. برای سنجش دقت الگوریتم‌ها نیز از شاخص‌های بهینه NMSE و MAE و DS استفاده نمودیم، هرچه NMSE و MAE کوچک‌تر باشند دقت پیش‌بینی بالاتر می‌باشد و شاخص DS هرچه بزرگ‌تر باشد، پیش‌بینی دقت بالاتری خواهد داشت. باتوجه‌به نتایج به‌دست‌آمده در جدول ۲ می‌توان گفت که مقادیر شاخص‌ها در الگوریتم ترکیبی GA-SVM بسیار مناسب‌تر الگوریتم SVM می‌باشند. آزمایش‌های پیشنهادی نشان می‌دهد که الگوریتم ترکیبی GA-SVM یک ابزار جایگزین امیدوارکننده برای ماشین‌های بردار پشتیبانی برای پیش‌بینی سری زمانی مالی فراهم می‌کند، زیرا اصل اصول کوچک‌سازی خطر ساختاری را تصویب می‌کند، در نهایت منجر به تعمیم بهتری نسبت به تکنیک معمولی می‌شود.

برای کارهای بعدی، پیشنهاد می‌شود که دررابطه‌با ترکیب ابزارها و مدل‌های مختلف پیش‌بینی قیمت سهام پژوهش شود. همچنین می‌توان برای غلبه کردن بر محدودیت‌های روش‌های کمی و پیش‌بینی بهتر، نتایج حاصل از روش‌های کیفی با روش‌های کمی را ترکیب کنیم. از الگوریتم‌های دیگر مانند KNN و ANN یا الگوریتم LSMT نیز می‌توان استفاده کرد. با حجم داده‌های کم و پس از آن برای داده‌های بیشتر آن را توسعه داد و تأثیرگذاری الگوریتم‌ها را سنجید. همچنین می‌توان ابتدا به بهینه‌سازی سبد سهام پرداخته و با استفاده از نتایج به‌دست‌آمده برای سبد سهام بهینه، به پیش‌بینی قیمت آتی و تغییرات بعدی سبد پرداخت.

٦- منابع

اسکندری، علی؛ نوری، روح‌اله؛ معراجی، حامد؛ کیاقادی، امین. (۱۳۹۱). توسعهٔ مدلی مناسب بر مبنای شبکهٔ عصبی مصنوعی و ماشین بردار پشتیبان برای پیش‌بینی بهنگام اکسیژن‌خواهی بیوشیمیایی ۵ روزه. *محیط شناسی، ۳۸*(۱)، ۸۲-۷۱.

البرزی، محمود(۱۳۹۳)، الگوریتم ژنتیک، تهران، *دانشگاه صنعتی شریف. موسسه انتشارات علمی.*

آذرافزا، هاله؛ رضایی، حسین؛ بهمنش، جواد؛ بشارت، سینا. (۱۳۹۱). مقایسه نتایج بکارگیری الگوریتم‌های PSO، GA و SA در بهینه‌سازی سیستم‌های تک‌مخزنه (مطالعه‌موردی: سدشهرچای، ارومیه). *آب و خاک، ۲۶*(۵)، ۱۱۰۸-۱۱۰۱.

دایی، امیر؛ عبادتی، امیدمهدی؛ برنا، کیوان. (۱۳۹۸). به کارگیری وب کاوی در پیش‌بینی جهت قیمت سهام گروه محصولات شیمیایی در بورس اوراق بهادار. *فصلنامه فناوری اطلاعات و ارتباطات ایران*، ۳۹(۱۱)، ۴۸-۱۹.

دولو، مریم؛ صفری، علی. (۱۳۹۵). پیش بینی شاخص قیمت سهام با استفاده از مدل هیبریدی. *فصلنامه بورس اوراق بهادار*، ۹(۳۵)، ۱۰۱-۸۱.

فلاح پور، سعید؛ گل ارضی، غلامحسین؛ فتوره چیان، ناصر. (۱۳۹۲). پیش‌بینی روند حرکتی قیمت سهام با استفاده از ماشین بردار پشتیبان برپایهٔ الگوریتم ژنتیک در بورس اوراق بهادار تهران. *تحقیقات مالی*، ۱۵(۲)، ۲۸۸-۲۶۹.

فهمی حسن، آرش؛ مغاری، محمدرضا؛ عبادتی، امیدمهدی. (۱۳۹۷). پیش بینی اهداء خون با استفاده از داده کاوی بر پایه الگوریتم های درخت تصمیم، KNN، SVM و MLP. *مدیریت مهندسی و رایانش نرم*، ۴(۱)، ۹۷-۷۷.

Alborzi, Mahmoud (2014), Genetic Algorithm, Tehran, *Sharif University of Technology. Scientific Publishing Institute.* [In Persian].

Azarafza, H., Rezaei, H., Behmanesh, J., Besharat, S. (2012). Results Comparison of Employing PSO, GA and SA Algorithms in Optimizing Reservoir Operation (Case Study: Shaharchai Dam, Urmia, Iran). *Water and Soil*, *26*(5), 1101-1108. [In Persian].

Cervantes, J., García Lamont, F., López-Chau, A., Rodríguez Mazahua, L., & Sergio Ruíz, J. (2015). Data selection based on decision tree for SVM classification on large data sets. *Applied Soft Computing*, *37*, 787–798.

Daee, A., Ebadati, O,M., Borna, K. (2018).Using web analytics in forecasting the stock price of chemical products group in the stock exchange. *Journal of Information and Communication Technology*. *39*(11) .19-48. [In Persian].

Davalou, M., Safari, A. (2016). Stock Price Index Forecasting Using a Hybrid Model. *Journal of Securities Exchange*, *9*(35), 81-101. [In Persian].

Ebadati E., O. M., & Mortazavi T., M. (2018). An Efficient Hybrid Machine Learning Method For Time Series Stock Market Forecasting. *Neural Network World*, *28*(1), 41–55.

Eskandari, A., Nouri, R., Meraji, H., Kiaghadi, A. (2012). Developing a Proper Model for Online Estimation of the 5-Day Biochemical Oxygen Demand Based on Artificial Neural Network and Support Vector Machine. *Journal of Environmental Studies*, *38*(1), 71-82. [In Persian].

Fahmi Hassan, A., Moghari, M., Ebadati, O. (2018). Prediction of Blood Donations Using Data Mining Based on the Decision Tree Algorithms KNN, SVM, and MLP. *Engineering Management and Soft Computing*, *4*(1), 77-97. [In Persian].

Fallahpour, S., Golarzi, G., Fatourechian, N. (2013). Predicting Stock Price Movement Using Support Vector Machine Based on Genetic Algorithm in Tehran Stock Exchange Market. *Financial Research Journal*, *15*(2), 269-288. [In Persian].

Fletcher, R (1987). Practical Methods of Optimization. *Wiley, New York.* 436.

Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Systems with Applications, 129*, 273–285.

Iqbal, Z., Ilyas, R., Shahzad, W., Mahmood, Z., & Anjum, J. (2013). Efficient machine learning techniques for stock market prediction. *International Journal of Engineering Research and Applications, 3*(6), 855-867.

Jena, O.P. and S. Padhy (2014).Application of GA with SVM for Stock Price Prediction in Financial Market. *International Journal of Science and Research (IJSR). 3*(10), 498-503.

Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M. (1990). Stock market prediction system with modular neural networks. *1990 IJCNN International Joint Conference on Neural Networks.*

Lamba, A., & Kumar, D. (2016). Survey on KNN and its variants. International *Journal of Advanced Research in Computer and Communication Engineering, 5*(5), 430-435.

Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management*, 102212.

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E., & S., S. (2020). Deep Learning for Stock Market Prediction. *Entropy, 22*(8), 840.

Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2018). An innovative neural network approach for stock market prediction. *The Journal of Supercomputing.*

Qiu, M., & Song, Y. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. *PLOS ONE, 11*(5), e0155133.

Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications, 79*, 153–163.

Zhang, K., Zhong, G., Dong, J., Wang, S., & Wang, Y. (2019). Stock Market Prediction Based on Generative Adversarial Network. *Procedia Computer Science, 147*, 400–406.

Forecasting Stocks in the Financial Market by Using GA-SVM Hybrid Algorithm

Omid Mahdi Ebadati[1], Mohammd Ali Jafari[2], Nasim Davoodifar[3]

Abstract

The purpose of this paper is to predict stock prices using Hybrid GA-SVM Algorithm. Predicting time series such as stock price forecasting is one of the most important issues in financial field. In real life, identifying time series movements in stock price indices is very complex. Therefore, the use of a classical model alone cannot accurately predict stock price indices. Hence, by using combined methods, uncertainty in forecasting can be reduced. In stock price forecasting in financial sector, more than 100 indicators have been created to understand stock market behavior, so, identifying the appropriate indicators is a challenging problem. One of the techniques that has recently been studied for serial forecasting is support regression Vector (SVR) or machine support vector (SVM). This study uses the GA-SVM hybrid algorithm to predict the stock price index. Experimental results show that Hybrid GA-SVM Algorithm provides a more appropriate and promising alternative to stock market forecasting.

Keywords: Forecast, Stock Price, GA Algorithm, SVM Algorithm, International Financial Markets.

JEL Classification: C6, D4, F3, G2, G4.

1. Department of Operation Management and Information Technology, Kharazmi University, Tehran, Iran. (Corresponding Author). ebadati@khu.ac.ir
2. Department of Financial Mathematics, Kharazmi University, Tehran, Iran.
3. Department of Financial Mathematics, Kharazmi University, Tehran, Iran.