Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 8, Number 3, Fall 2019, pp. 60-67
Received: 02/18/2019, Revised: 06/25/2019, Accepted: 07/23/2019

Archive of SID

Machine learning algorithms for time series in financial markets

Maryam Dehghani[1], Mohammad Ghasemzadeh[2*], Habib Ansari-samani[3]
1,2*- Computer Engineering Department, Faculty of Engineering, Yazd University, Yazd, Iran.
3- Economics, Management & Accounting Department, Faculty of Economics, Yazd University, Yazd, Iran.
[1] mdehghani26@stu.yazd.ac.ir, [2*] m.ghasemzadeh@yazd.ac.ir, [3] h.samani@yazd.ac.ir

Corresponding author's address: Mohammad Ghasemzadeh, Faculty of Engineering, Computer Engineering Group, Yazd University, Yazd, Iran.

Abstract- This research focuses on the usefulness of various intelligent machine learning algorithms on prediction of time series in financial markets. A challenge in this area is that economic managers and the scientific community are still demanding predictive algorithms with greater accuracy. The elimination of the mentioned challenge can improve the quality of the predictions and, as a result, lead to higher profitability and productivity. The proposed solution relies on finding the best input variables by using the regression-based machine learning algorithms, with emphasis on the leading selection methods. We implemented the concerned ideas using the Python language and the relevant machine learning tools. In our experiments, as dataset, we used the stock information of two companies from the Tehran Stock Exchange. These datasets belong to the transactions accomplished in years 2008 to 2018. The experimental results show that the technical features selected by the forward method can find the most effective and also the best values for the required parameters. The experimental results and formal analyses indicate that the use of selected technical features as inputs to the support-vector-machine and to the multi-layer perceptron machine gives prediction with the least-error, and this would provide more accurate predictions.

Keywords- Prediction, Time Series, Machine Learning, Financial Markets, Stock Market.

www.SID.ir

مجله علمی پژوهشی رایانش نرم و فن‌آوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۸، شماره ۳، پاییز ۱۳۹۸، صفحه ۶۷-۶۰
دریافت: ۱۳۹۷/۱۱/۲۹، بازنگری: ۱۳۹۸/۰۴/۰۴، پذیرش: ۱۳۹۸/۰۵/۰۱

الگوریتم‌های یادگیری ماشین برای سری‌های زمانی در بازارهای مالی

مریم دهقانی[۱]، محمد قاسم زاده *[۲]، حبیب انصاری سامانی[۳]
۱- پردیس فنی و مهندسی، گروه مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران.
*۲- پردیس فنی و مهندسی، گروه مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران.
۳- دانشکده اقتصاد، مدیریت و حسابداری، گروه اقتصاد، دانشگاه یزد، یزد، ایران.
[1] mdehghani26@stu.yazd.ac.ir, [2*] m.ghasemzadeh@yazd.ac.ir , [3] h.samani@yazd.ac.ir

* نشانی نویسنده مسئول: محمد قاسم زاده، یزد، دانشگاه یزد ، پردیس فنی و مهندسی، گروه مهندسی کامپیوتر.

چکیده- این پژوهش در رابطه با بررسی سودمندی الگوریتم‌های هوشمند مختلف در حوزه یادگیری ماشین برای پیش‌بینی سری‌های زمانی در بازارهای مالی می‌باشد. چالش مورد توجه در این حوزه، این است که مدیران اقتصادی و جامعه علمی، همچنان خواستار مدل‌های پیش‌بینی با دقت بیشتری می‌باشند. رفع چالش یاد شده موجب ارتقای کیفیت پیش‌بینی و به جهت آن، سودآوری و بهره‌وری بالاتری می‌شود. راه حل پیشنهادی، تکیه بر بکارگیری الگوریتم‌های یادگیری ماشینِ مبتنی بر رگرسیون، با تأکید بر روش انتخاب ویژگی پیشرو، جهت یافتن بهترین متغیرهای فنی ورودی دارد. موارد یاد شده، با بکارگیری ابزارهای یادگیری ماشین به زبان پایتون پیاده‌سازی گردیدند. داده‌های تحقیق که در این پژوهش به کار گرفته شدند، اطلاعات مربوط به سهام دو شرکت از بورس تهران می‌باشند. این داده‌ها مربوط به سال‌های ۱۳۸۷ تا ابتدای سال ۱۳۹۷ می‌باشند. نتایج آزمایشی نشان می‌دهند که ویژگی‌های فنی منتخب توسط روش پیشرو، مؤثرترین و نیز بهترین مقادیر برای پارامترهای الگوریتم‌های یادگیری مورد نظر را می‌یابند. نتایج آزمایشی و تحلیل‌های رسمی دلالت بر این دارند که بکارگیری ویژگی‌های فنی منتخب ،به‌عنوان ورودی دو الگوریتم ماشین بردار پشتیبان و ماشین پرسپترون چند لایه، یک پیش‌بینی با حداقل خطا را در اختیار می‌گذارد. این مطلب موجب پیش‌بینی با دقت بالاتری می‌گردد.

واژه‌های کلیدی: پیش‌بینی، سری‌های زمانی، یادگیری ماشین، بازارهای مالی، بازار سهام.

۱- مقدمه

یکی از مسائل مهم در زمینه‌ی سری‌های زمانی، پیش‌بینی سری‌های زمانی مالی است که توجه بسیاری را به خود جلب کرده است. پیش‌بینی سری‌های زمانی مالی به مسائل سرمایه گذاری و تصمیم‌گیری‌های اقتصادی پرداخته و مدلی سودآور را در آینده بر اساس نتایج اقتصادی گذشته ارائه می‌دهد. پیش‌بینی سری‌های زمانی مالی مسئله‌ای چالش‌برانگیز در حوزه سری‌های زمانی محسوب می‌شود و نظر بسیاری از پژوهشگران را به خود جلب نموده است[۱]. هم اکنون یکی از اهداف اصلی گردانندگان بازارهای پولی و مالی این است که هر کسی، با هر سلیقه و هر مقدار و انتخاب هر نوع دارایی بتواند وارد این بازارها شده، فرصت-های مناسب سرمایه‌گذاری را تشخیص دهد و در صورت تشخیص صحیح بتواند سود مناسبی کسب نماید.

دنیای امروز دنیای تغییر است و دانستن این‌که در آینده چه موقعیتی در انتظار است، عامل مهمی در حفظ و بقای سازمان‌ها می‌باشد. عاملان اقتصادی هر روز با موقعیت‌ها و شرایطی روبرو می‌شوند که نیازمند پیش‌بینی آینده است. دولت به پیش‌بینی واردات و صادرات، سهامداران به شناخت وضعیت بازار و مدیران سازمان‌ها به شناخت رفتار کاری کارکنان نیاز دارند. مدیران هر روز تصمیمات شخصی و حرفه‌ای می‌گیرند که مبتنی بر پیش‌بینی

www.SID.ir

Archive of SID

وضع آینده است. در بسیاری از موارد پیش‌بینی آینده بر مبنای گذشته و حال است. درواقع ایشان سعی می‌کنند بین دو یا چند متغیر به نحوی ارتباط برقرار کنند که بتوانند از آن در پیش‌بینی آینده استفاده نمایند.

بازارهای مالی نقش مهمی در سازمان‌دهی اقتصادی و اجتماعی جامعه مدرن ایفا می‌کنند. بی‌تردید امروزه بیشترین مقدار سرمایه از طریق بازارهای بورس در تمام جهان مبادله می‌شود. اقتصادهای ملی به‌شدت متأثر از عملکرد پیش‌بینی قیمت سهام در بازار بورس است. یکی از مهم‌ترین اطلاعات در بازار بورس اوراق بهادار برای سرمایه‌گذاران، اطلاعات قیمت سهام است. بنابراین پیش‌بینی قیمت سهام نه فقط خیلی چالش‌انگیز هست بلکه موردعلاقه زیاد سرمایه‌گذاران می‌باشد. از طرفی سرمایه‌گذاری در بورس بخش مهمی از اقتصاد را تشکیل می‌دهد، لذا موضوع پیش‌بینی به‌ویژه در کشورهای در حال توسعه نظیر کشور ما، جهت مدیریت صحیح بورس اوراق بهادار برای رسیدن به توسعه پایدار اهمیت به سزایی را داراست. که تصمیم‌گیری در شرایط نامطمئن را برای تصمیم-گیران اجرایی بورس هموار می‌نماید از طرف دیگر سرمایه‌گذاران نیز می‌توانند قیمت سهام و یا شاخص کل را پیش‌بینی نموده و نسبت به آن تصمیمات منطقی اتخاذ نمایند.

با توجه به اهمیت موضوع بر اساس مطالعات انجام شده الگوریتم-های یادگیری ماشین عملکرد چشمگیری در این زمینه داشته‌اند لذا امروزه اکثر پژوهش‌ها در زمینه پیش‌بینی قیمت سهام با محوریت روش‌های هوشمند می‌باشد[۲]. ازاین‌رو انجام پژوهش بیشتر در این زمینه امری ضروری، به نظر می‌رسد. لذا روش‌های یادگیری ماشین به دلیل توانمندی بالایی که در مدل کردن مسائل پیچیده و سیستم‌های غیرخطی دارند برای مسئله پیش-بینی قیمت سهام به‌عنوان روش‌هایی سودمند شناخته شده‌اند.

مهم‌ترین هدفی که در این پژوهش مد نظر است افزایش دقت پیش‌بینی قیمت ماهانه سهام بازار بورس ایران به‌عنوان یک سری زمانی مالی با به‌کارگیری الگوریتم‌های یادگیری و ویژگی‌های تکنیکال و بنیادی است. به‌منظور دستیابی به هدف فوق، چگونگی انتخاب مناسب‌ترین ویژگی‌ها و تنظیم پارامترهای اصلی وتأثیرگذار در الگوریتم‌های انتخاب شده حائز اهمیت است.

در این پژوهش نشان داده می‌شود که چگونه می‌توان با بهره بردن از تحلیل فنی سری‌های زمانی مالی موجود در داده‌های بازار بورس، ضمن دسترسی به مؤثرترین ویژگی‌ها، به‌دقت مطلوب دست یافت. اگرچه در مطالعات و پژوهش‌های قبلی از تحلیل فنی برای کشف ویژگی‌های فنی بهره برده‌اند اما در پژوهش حاضر در نظر گرفتن تعداد بیشتری از ویژگی‌های فنی، تنوع و کنار هم قرار گرفتن این نوع از ویژگی‌ها یک نوآوری تلقی می‌گردد.

نوآوری‌های ارائه شده در این پژوهش به شرح زیر می‌باشد:

۱- دادگان انتخابی، دادگانی می‌باشند که برای اولین بار با ویژگی‌های پیشنهادی مورد ارزیابی قرار می‌گیرند.

۲- به‌کارگیری پرکاربردترین الگوریتم‌های یادگیری مبتنی بر رگرسیون برای پیش‌بینی قیمت سهام مختلف.

۳- انتخاب ۲۳ ویژگی پیشنهادی بر پایه‌ی تحلیل فنی محاسبه‌شده.

۲- دانش پس‌زمینه

در این بخش ابتدا مفاهیم مرتبط با سری زمانی مالی و همچنین یادگیری ماشین ارائه می‌شوند. در ادامه، پس از آشنایی با بازار بورس تهران، چگونگی پیش‌بینی سری‌های مالی توسط الگوریتم-های یادگیری بیان می‌شود و سپس به بررسی روش‌های موجود در پیش‌بینی سری‌های زمانی مالی پرداخته می شود.

۲-۱- سری زمانی مالی

سری زمانی مالی اغلب شامل قیمت اقلام، سهام و برخی معیارهای مرتبط همچون حجم سهام مبادله شده است. یکی از مسائل مهم در زمینه‌ی سری‌های زمانی، پیش‌بینی سری‌های زمانی مالی است که توجه بسیاری را به خود جلب کرده است .پیش‌بینی سری‌های زمانی مالی به مسائل سرمایه‌گذاری و تصمیم‌گیری‌های اقتصادی پرداخته و یک مدل اقتصادی سودآور را در آینده بر اساس نتایج اقتصادی گذشته ارائه می‌دهد.

۲-۲- بازارهای مالی

بخش مالی را می‌توان به عنوان زیرمجموعه‌ای از نظام اقتصادی تعریف کرد که در آن وجوه، اعتبارات و سرمایه در چارچوب قوانین و مقررات مشخص از طرف پس‌انداز کنندگان و صاحبان پول و سرمایه به طرف متقاضیان، جریان می‌یابد. بازارهای مالی نیز بازارهایی هستند که در آنها دارایی‌های مالی مبادله می‌شوند. دارایی‌های مالی دارایی‌هایی مثل سهام و اوراق قرضه هستند که ارزش آنها به ارزش تولیدات و خدمات ارائه شده توسط شرکت‌های منتشرکننده آنها وابسته است. تفاوت دارایی‌های مالی با دارایی-های واقعی در این است که دارایی‌های واقعی ماهیت فیزیکی دارند مانند اتومبیل، املاک و مستغلات، و وسایل منزل[۳].

در اقتصاد، بازار مالی به مکانیسمی اطلاق می‌شود که امکان خریدوفروش سهام، کالا یا هر محصول قابل‌تبادلی را برای

www.SID.ir

Archive of SID

مشارکت‌کنندگان با هزینه‌ای پایین فراهم می‌آورد. بازارهای مالی چند سال اخیر رشد قابل‌توجهی داشته و به‌طور مستمر در حال تکامل و افزایش نقدینگی در جهان است. وظیفه بازارهای مالی گردآوری خریداران و فروشندگان علاقه‌مند به مشارکت در بازار است. درواقع تمامی بازارهای مالی در پی گردآوری سرمایه و برقراری ارتباط بین جویندگان سرمایه (قرض گیرندگان) و دارندگان سرمایه (قرض دهندگان) هستند.

۲-۳- بازار بورس تهران

برای هـر کالایـی معمولاً یک بازار وجود دارد، بازار خودرو، بازار سیمان، بازار فرش، بازار طلا. بعضی از بازارها سنتی و خاطره‌انگیز هستند، بعضی هم پیشرفتـه و امروزی. اما همه بازارها، یک ویژگی مشتـرک دارند آن‌هم این‌که در آنها، کالاها یا به عبارت دیگر، دارایی‌های مختلف خریدوفروش می‌شود، در تعریف کلی، بازار، محل خریدوفروش دارایـی است. اما دارایی چه انواعی دارد. در یک طبقه‌بندی کلی، دارایی به دو نوع دارایی واقعی و دارایی مالی تقسیم‌بندی می‌شود. بورس اوراق بهادار تهران از پانزدهم بهمن ۱۳۴۶، فعالیت خود را به‌طور رسمی با پذیرش سهام بانک توسعه صنعتی و معدنی ایران به‌عنوان بزرگ‌ترین مجتمع واحدهای تولیدی و اقتصادی آن زمان و سپس سهام شرکت نفت پارس، اوراق قرضه دولتی، اسناد خزانه و اوراق قرضه عباس‌آباد آغاز کرد. هر داده سهام شامل اطلاعاتی از قبیل تاریخ، اولین قیمت، بیشترین قیمت، کمترین قیمت، قیمت پایانی، حجم، ارزش، تعداد معاملات، قیمت روز گذشته، سهام در آن روز است.

۲-۴- پیش‌بینی در سری زمانی مالی

وقتی بحث پیش‌بینی در بازارهای مالی مطرح می‌شود ابتدا باید به تجزیه‌وتحلیل داده‌ها پرداخت[۴]. از دو روش معمول برای تجزیه‌وتحلیل در سری‌های مالی استفاده می‌شود: تجزیه و تحلیل فنی، تجزیه‌وتحلیل بنیادی (پایه‌ای).

تجزیه‌وتحلیل بنیادی

نوعی روش تجزیه‌وتحلیل است که بر اطلاعات و آمار کلیدی مندرج در صورت‌های مالی شرکت‌ها، آمار اقتصاد کلان کشور و عوامل تأثیرگذار بر صنایع مختلف تمرکز دارد؛ لذا در آن خریدوفروش سهام یا هر نوع دارایی، پس از بررسی اطلاعات کامل در خصوص عوامل ذکرشده، صورت می‌گیرد[۵] .

تجزیه‌وتحلیل فنی

روشی برای پیش‌بینی قیمت در بازارها، بر اساس مطالعه و بررسی تاریخچه قیمت و حجم معاملات در آن بازار است. تحلیل فنی، که فرمول‌های ریاضی مربوط به قیمت یا اطلاعات حجم هستند، برای مدل‌سازی برخی از جنبه‌های قیمت سهام یا شاخص‌های سهام، مورداستفاده قرار می‌گیرند[۶] . درواقع تحلیل فنی شامل مطالعه و بررسی شاخص‌ها، نمودارها و الگوهای مختلفی است که روند بازار و وضعیت سهام مختلف را برای سرمایه‌گذار روشن می‌کند.

۲-۵- روش‌های یادگیری ماشین

روش‌های یادگیری ماشین، که از الگو شناسی و نظریه یادگیری محاسباتی الهام گرفته‌شده است، مطالعه و ساخت الگوریتم‌هایی را که می‌توانند بر اساس داده‌ها یادگیری و پیش‌بینی انجام دهند بررسی می‌کند. چنین الگوریتم‌هایی از دستورات برنامه پیروی نمی‌کنند و از طریق مدل‌سازی از داده‌های ورودی نمونه، پیش-بینی یا تصمیم‌گیری را انجام می‌دهند.

این روش‌ها در کارهای محاسباتی که طراحی و برنامه‌نویسی الگوریتم‌های صریح با عملکرد مناسب در آن‌ها سخت یا نشدنی است، استفاده می‌شود، که با یادگیری از داده‌ها، قواعد استخراج می‌شود تا بر اساس آن‌ها بتوان نرم‌افزاری ایجاد کرد تا نیازهای کاربران را مرتفع سازد. بنابراین در یادگیری ماشین امکانی برای توسعه نرم‌افزار فراهم می‌آید که بخشی از قواعد سیستم نرم‌افزاری که به‌واسطه تحلیل و طراحی توسط تحلیلگر و طراح قابل‌استخراج نیست از داده‌های گذشته به دست آید. عملاً در حوزه یادگیری به این مجموعه قواعد استخراج شده مدل گفته می‌شود. ازاین‌رو می‌خواهیم مدل استخراج شده تا حد امکان عمومی باشد تا بتوان از آن برای داده‌هایی که هنوز ایجاد نشده‌اند با دقت بالا بهره برد.

۳- پیشینه پژوهش

پیش‌بینی سری زمانی مالی می‌تواند یکی از چالش‌های اصلی در حوزه سری‌های زمانی و یادگیری ماشین باشد. در دهه‌های گذشته، برای پیش‌بینی بازارهای مالی و ارائه سیستم‌های پشتیبانی تصمیم‌گیری، چندین روش پیشنهادشده است. تکنیک-های محاسبات نرم مانند سیستم‌های خبره، سیستم‌های فازی و شبکه‌های عصبی با موفقیت نسبی در مدل‌سازی و پیش‌بینی سری‌های مالی به کار گرفته‌شده‌اند [۷]. در مقایسه با روش‌های پیش‌بینی آماری سنتی، تکنیک‌های محاسبات نرم قادر به گرفتن رابطه غیرخطی بدون داشتن هیچ دانش پیش‌زمینه درباره توزیع داده‌های ورودی هستند.

اخیراً، شبکه‌های عصبی مصنوعی در زمینه پیش‌بینی بازار مالی بسیار محبوب شده‌اند. شبکه‌های عصبی مصنوعی روش‌های مبتنی

www.SID.ir

Archive of SID

برداده و خودسازگاری هستند که توانایی این را دارند که رفتارهای غیرخطی سری زمانی را بدون هیچ‌گونه فرضیه آماری در مورد داده‌ها شناسایی کنند. به عنوان مثال هو و همکارانش[۸] به این نتیجه رسیدند که شبکه‌های عصبی مصنوعی بهتر از روش‌های آماری کلاسیک مثل رگرسیون خطی و روش‌های باکس-جنکینز عمل می‌کند. یک مطالعه مشابه، توسط پانگ و همکارانش[۹] به عمل آمد که نشان می‌داد از شبکه‌های عصبی مصنوعی می‌توان با موفقیت برای مدل‌سازی و پیش‌بینی سری‌های زمانی غیرخطی استفاده کرد.

چندین نوع شبکه عصبی پیشنهاد شده که در پژوهش‌ها برای پیش‌بینی بازار مالی پرکاربرد هستند روش‌های شبکه عصبیِ پرسپترون تک لایه، پرسپترون چندلایه می‌باشد که شبکه عصبیِ چندلایه باوجود پیچیدگیِ بیشترِ آن، بیشتر از دو روش دیگر در مدل‌سازی‌ها بکار رفته است. در اغلب پژوهش‌ها از MLP[1] برای یادگیری رابطه بین برخی از شاخص‌های فنی و پیش‌بینی حداکثر و حداقل قیمت سهام روزانه استفاده شده است.

به عنوان مثال لسمان و همکارانش[۱۰] برای پیش‌بینی آینده قیمت سهام بورس اوراق بهادار بنگلادش الگوریتم MLP را با ترکیب دو ویژگی فنی میانگین متحرک همگرایی-واگرایی و شاخص توان نسبی مورد بررسی قرار دادند. از الگوریتم MLP برای یادگیری برخی از مقادیر مالی و ترازنامه‌های شرکت‌ها مطابق با تحلیل‌های بنیادی استفاده کرده است تا شرکت‌ها را بر اساس مشکلات مالی طبقه‌بندی کنند.

کاوالکانت و همکارانش[۱۱] الگوریتم‌های یادگیری مختلف را در زمینه پیش‌بینی بازارهای مالی ارائه می‌دهند. هرکدام از این الگوریتم‌ها پیش‌بینی را بسته به این‌که متغیرهای ورودی خود را با چه نوع تحلیل سری زمانی مالی دریافت کنند، انجام می‌دهند. رویکردهای پیشنهادی در پژوهش‌های مختلف تعداد و نوع متغیرهای مورد استفاده برای مدل‌سازی رفتار بازار مالی متفاوتند[۱۲] .

به‌عنوان مثال چن و هاو [۱۳] با تجزیه‌وتحلیل فنی مجموعه‌ای از ویژگی‌های فنی مانند میانگین متحرک و میانگین متحرک نمایی و حرکت متوسط همگرایی-واگرایی و نسبت حجم و شاخص توان نسبی و میزان تراز حجم و شاخص شتاب را برای پیش‌بینی با الگوریتم‌های یادگیری را ارائه می‌دهند.

باراک و مدرس[۱۴] با تجزیه‌وتحلیل بنیادی چندین ویژگی کلیدی و اقتصادی یک شرکت از قبیل اندازه شرکت، سرمایه، نسبت قیمت درآمد و جریان نقدی را ارائه می‌دهد.

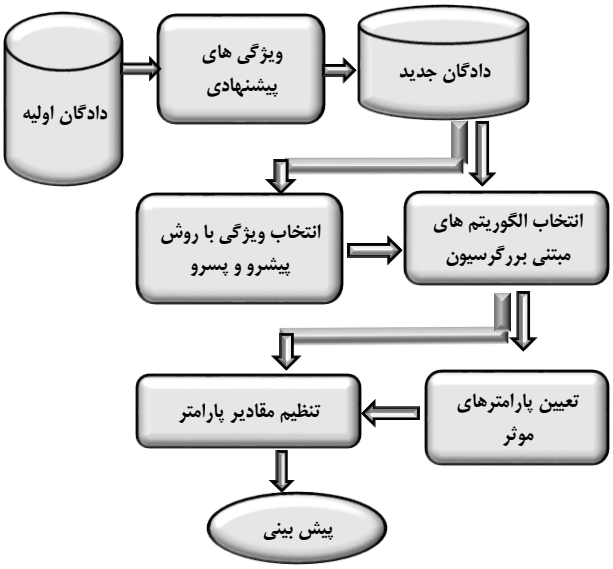
۴- سؤال پژوهش و راهبرد پیشنهادی

سؤال پژوهش عبارت است از اینکه: چگونه می‌توان دقت پیش‌بینی قیمت سهام بازار بورس ایران به عنوان یک سری زمانی مالی را با به‌کارگیری الگوریتم‌های یادگیری بصورت میان‌مدت (ماهانه) افزایش داد.

۴-۱- راهبرد پیشنهادی

در این پژوهش نشان می‌دهیم که چگونه می‌توان با بهره بردن از تحلیل فنی سری‌های زمانی مالی موجود در داده‌های بازار بورس، ضمن دسترسی به مؤثرترین ویژگی‌ها، به‌دقت مطلوب دست‌یافت.

در این پژوهش استفاده از روش‌های پیشرو و پسرو به دلیل در نظر گرفتن همه زیرمجموعه‌های ممکن از ویژگی‌ها می‌باشد. و همچنین با ارزیابی همه حالت‌های ممکن و امتیازدهی به ویژگی-ها، بهترین حالت‌ها با کمترین خطا انتخاب می‌شود. راه‌حل پیشنهادی، تکیه‌بر به‌کارگیری الگوریتم‌های یادگیری ماشینِ مبتنی بر رگرسیون، با تأکید بر روش انتخاب ویژگی پیشرو، جهت یافتن بهترین متغیرهای فنی ورودی دارد. روند کلی پژوهش در شکل (۱) قابل‌مشاهده است.



شکل۱: روند روش پیشنهادی

۴-۲- تشریح روش پیشنهادی

دادگان اولیه این تحقیق، اطلاعات سهام دو شرکت در بازار بورس تهران است که هر داده‌ی سهام شامل ده ویژگی پایه می‌باشد. این ویژگی‌ها که عبارتند از: تاریخ، اولین قیمت، بیشترین قیمت،

www.SID.ir

Archive of SID

کمترین قیمت، قیمت پایانی، حجم، ارزش، تعداد معاملات و قیمت روز گذشته سهام در هرروز.

با استفاده از تحلیل فنی، ۲۳ ویژگی فنی از روی اطلاعاتی که از ۱۰ ویژگی پایه برای سهام شرکت‌ها در بازه زمانی ده‌ساله جمع-آوری شده بود، محاسبه شد. استخراج ۲۳ ویژگی فنی و افزودن به دادگان اولیه با ده ویژگی پایه، دادگان جدید ایجاد می‌شود.

ویژگی‌های پیشنهادی (شاخص کل قیمت، شاخص صنعت، شاخص قیمت هم‌وزن، شاخص صنایع، بازده کل، بازده صنعت، ضریب بتا بازده صنعت، ضریب بتا بازده شاخص کل، میانگین متحرک همگرایی-واگرایی، میانگین متحرک سه‌روزه، میانگین متحرک پنج‌روزه، میانگین متحرک ده روزه، میانگین متحرک بیست روزه، میانگین متحرک سی روزه، میانگین متحرک هفت‌روزه حجم، میانگین متحرک وزنی، شاخص توان نسبی، باند بالایی بولینگر، باند پایینی بولینگر، اولین روزهای هر هفته، آخرین روزهای کاری بازار سهام، اولین ماه‌های هرسال و شاخص نرخ تغییر) که بر پایه تحلیل فنی می‌باشند.

بعد از جمع‌آوری ویژگی‌های پیشنهادی سه الگوریتم پرکاربرد مبتنی بر رگرسیون در حوزه پیش‌بینی بازارهای مالی، الگوریتم ماشین بردار پشتیبان، پرسپترون چندلایه و درخت تصمیم را انتخاب نمودیم. سه الگوریتم انتخاب‌شده نماینده‌ای از خانواده‌های الگوریتم‌های یادگیری می‌باشد، به‌عنوان‌مثال الگوریتم DT[2] از خانواده درخت‌ها، الگوریتم MLP از خانواده شبکه‌های عصبی و الگوریتم SVR[3] از خانواده بردارهای پشتیبان است. به عبارتی هر سه از الگوریتم‌های پایه و عمومی در یادگیری ماشین هستند. ویژگی‌های فنی با روش انتخاب ویژگی پیش‌رو و پس‌رو به‌عنوان متغیرهای ورودی در ۲۳ مرحله به الگوریتم‌ها داده می‌شوند.

۵- یافته‌های پژوهش

در پیش‌بینی سری‌های مالی با الگوریتم‌های یادگیری ماشین، آنچه اهمیت دارد، این است که با ارزیابی و مقایسه‌ی نتایج الگوریتم‌ها، بتوان به بهترین پیش‌بینی با میزان خطای حداقل دست یافت. ازجمله عوامل مؤثر در موفقیت الگوریتم‌ها، آن است که الگوریتم با معیارهای مناسبی موردسنجش قرار گیرد.

۵-۱- دادگان ارزیابی

داده‌های تحقیق که در این پژوهش به‌کاررفته است شامل سهام دو شرکت در بازار بورس تهران می‌باشند. که از ابتدای سال ۱۳۸۷ تا ابتدای سال ۱۳۹۷ از نرم‌افزار TseClient، به‌دست‌آمده است. داده‌ها شامل ۲۰۰۰ نمونه و ۱۰ ویژگی پایه: تاریخ، اولین قیمت، بیشترین قیمت، کمترین قیمت، قیمت پایانی، حجم، ارزش، تعداد معاملات، قیمت روز گذشته، سهام در آن روز است. بعد از جمع-آوری دادگان در اولین مرحله برای پاکسازی دادگان با استفاده از تابع Dropna از کتابخانه pandas موجود در پایتون داده‌های بدون مقدار از دادگان حذف شدند، و سپس دادگان نرمال‌سازی شدند. بر اساس بررسی‌های انجام شده، در کارهایی که در حوزه‌ی پیش‌بینی سری‌های مالی با الگوریتم‌های یادگیری ماشین انجام‌شده، دادگان ارزیابی را به‌واسطه‌ی روش‌های اعتبارسنجی به دودسته‌ی مجموعه آموزشی و مجموعه آزمون تقسیم نموده و از مجموعه آموزشی، برای آموزش الگوریتم‌ها استفاده می‌نمایند. البته شایان‌ذکر است که این داده‌ی آموزشی هم، خود به دو دسته‌ی آموزش و اعتبارسنج که برای ارزیابی میزان آموزشی که از دادگان آموزشی حاصل‌شده مورد استفاده قرار می‌گیرد تقسیم می‌شود. نتیجه‌ی این ارزیابی فقط به جهت انتخاب بهترین بخش آموزشی مورداستفاده قرار می‌گیرد. مسلم است که نتیجه‌ی حاصل از ارزیابی دادگانی که در آموزش بکار گرفته شدند، به‌عنوان نتیجه‌ی ارزیابی کلی مدنظر قرار نمی‌گیرد بلکه آنچه به‌عنوان نتیجه‌ی ارزیابی دقت الگوریتم مدنظر لحاظ می‌شوند.

برای مشخص کردن مجموعه آزمون و آزمایش برای هر سهم، ۶۰ درصد داده‌ها به‌عنوان مجموعه آموزش، ۲۰ درصد به‌عنوان مجموعه آزمون و۲۰ درصد هم به‌عنوان مجموعه اعتبارسنجی در نظر گرفته‌شده است که در جدول (۱) نشان داده‌شده است.

جدول ۱: تعداد نمونه‌های آموزش و آزمون برای هر ۲ سهم

نام سهم	تعداد کل نمونه هر سهم	مجموعه آموزش	مجموعه آزمون
شرکت نفت پارس	۶۱۶	۴۱۴	۱۴۰
پتروشیمی شازند	۷۱۶	۴۴۳	۱۳۸

در سری‌های زمانی مالی، اعتبار سنجی متقابل سنتی (مانند k-fold) به دلیل وابستگی‌های زمانی و انتخاب دلخواه از مجموعه آزمون قابل‌استفاده نیست.

به همین منظور مراحل روش اعتبارسنجی که در سری‌های زمانی مالی مورداستفاده قرار می‌گیرد، بدین شرح است.

مرحله اول؛ کل داده به چند بخش تقسیم می‌شود. مقدار پیش فرض این تقسیم‌بندی در پیاده‌سازی‌های مرسوم مقدار سه است

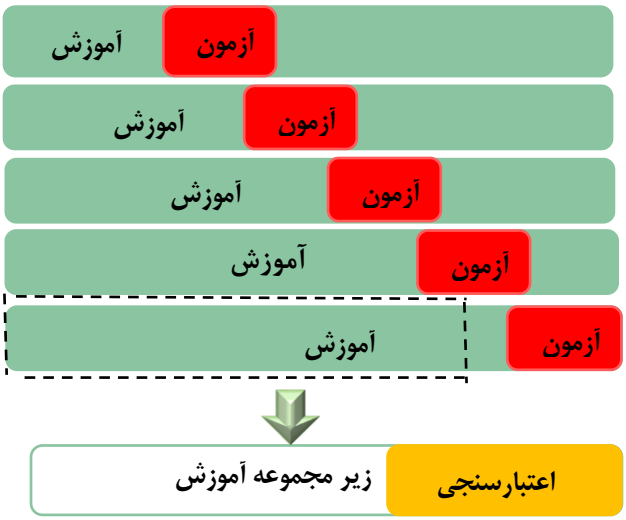
www.SID.ir

Archive of SID

که در این تحقیق با چند بار آزمایش مقدار ده برای تقسیم‌بندی داده‌ها مقداری مناسب در نظر گرفته شد. درنتیجه داده‌ها ده قسمت می‌شود و در هر مرحله یک قسمت انتخاب می‌شود.

مرحله دوم؛ بخش منتخب از ۱۰ قسمت به مجموعه آزمون و آموزش تقسیم می‌شوند و اعتبارسنجی انجام می‌شود.

مرحله سوم؛ در این مرحله مجدداً قسمتی دیگر از ۱۰ قسمت انتخاب‌شده و قسمتی که در مرحله دوم، مورد اعتبارسنجی قرار گرفت به این قسمت اضافه می‌شود. مجموعه آموزش و آزمون را مشخص کرده و اعتبارسنجی انجام می‌دهیم. این روند همان‌طور که در شکل(۲) قابل‌مشاهده است تا وقتی که هر ۱۰ قسمت مورد ارزیابی قرار گیرند، ادامه می‌یابد.



شکل ۲: اعتبارسنجی داده‌های سری زمانی مالی

برای انجام ارزیابی‌ها ابتدا یک‌بار دادگان اولیه با ده ویژگی پایه مورد ارزیابی قرار می‌گیرند، سپس ارزیابی به روی دادگان جدید انجام می‌شود. دادگان جدید، دادگانی با ویژگی‌های پیشنهادی هستند که از روش پیشرو و پس‌رو، انتخاب‌شده و به‌عنوان ورودی به الگوریتم‌ها داده می‌شوند.

۲-۵- معیارهای ارزیابی

برای ارزیابی الگوریتم‌های یادگیری با روش پیشنهادی از معیارهای استاندارد در حوزه پیش‌بینی سری‌های مالی، خطای میانگین مربعات و خطای جذر میانگین مربعات استفاده شده است. این معیارها به‌صورت گسترده‌ای در پیش‌بینی استفاده می‌شود.

خطای میانگین مربعات؛ همان‌طور که از نام این مفهوم مشخص است، سه عمل روی‌داده انجام می‌شود:

۱) محاسبه خطای خروجی الگوریتم، ۲) به توان دو رساندن خطای الگوریتم و ۳) میانگین‌گیری از مجموع تمام خطاها که هرکدام به توان دو رسیده است و از رابطه ۱ قابل‌محاسبه است:

$$MSE = \frac{1}{N}\sum(f_i - y_i)^2 \qquad (۱)$$

مقادیر هر متغیر در رابطه ۱ عبارت‌اند از، f_i معادل خروجی الگوریتم و y_i معادل پاسخ صحیح(قطعی) می‌باشد.

خطای جذر میانگین مربعات؛ خطای جذر میانگین مربعات، یک معیار خطای بسیار پرکاربرد برای اندازه‌گیری دقت برای زمان پیوسته است که در رابطه (۲) تعریف‌شده است.

$$RMSE = \sqrt{\frac{1}{N}\sum(f_i - y_i)^2} \qquad (۲)$$

همان‌طور که در جدول (۲) قابل‌مشاهده است، اعتبارسنجی و آزمون برای سهام شرکت نفت پارس، با دادگان اولیه ، الگوریتم SVR عملکرد بهتری نسبت به MLP و DT با هردو معیار ارزیابی MSE و RMSE داشته است.

جدول ۲: نتایج ارزیابی سهام شرکت نفت پارس با دادگان اولیه

الگوریتم	RMSE(اعتبارسنجی)	MSE(اعتبارسنجی)
SVR	۰/۰۶۱۸۰	۰/۰۰۳۸۲
MLP	۰/۰۶۹۷۸	۰/۰۰۴۸۷
DT	۰/۰۷۶۴۱	۰/۰۰۵۸۴
الگوریتم	**RMSE(آزمون)**	**MSE(آزمون)**
SVR	۰/۲۳۱۶۱	۰/۰۵۳۶۴
MLP	۰/۱۹۶۴۰	۰/۰۳۸۵۷
DT	۰/۲۰۳۴۰	۰/۰۴۱۳۷

با توجه به نتایج به‌دست‌آمده در جدول (۳) برای همین سهم هر سه الگوریتم با ویژگی‌های پیشنهادی نسبت به دادگان اولیه، پیش‌بینی را با میزان خطای کمتری انجام داده‌اند. درواقع هم در اعتبارسنجی و هم آزمون هر سه الگوریتم با ویژگی‌های پیشنهادی عملکرد بهتری نسبت به پیش‌بینی الگوریتم‌ها در دادگان اولیه داشته‌اند. نتایج معیارهای ارزیابی در جدول (۳) نشان می‌دهد برخلاف اینکه با دادگان اولیه الگوریتم SVR نسبت به دو الگوریتم دیگر عملکرد بهتری داشت اما الگوریتم MLP با ویژگی‌های پیشنهادی شاخص صنعت، بازده صنعت، میانگین متحرک ۳۰ روزه، بتای صنعت و بازده کل نسبت به دو الگوریتم دیگر عملکرد

www.SID.ir

Archive of SID

بهتری داشته است. علاوه بر این در اعتبارسنجی، برای هر سهم، مرحله یک تا پنج انتخاب ویژگی Forward به‌عنوان بهترین ویژگی‌ها با کمترین میزان خطا مشخص شد. (با توجه به این نکته که هرکدام از مراحل یک تا پنج حالات مختلفی دارند.)

همان‌طور که در بالا شرح داده شد نتایج اعتبارسنجی، برای سهم پارس با الگوریتم SVR در مرحله یک تا پنج معیارهای ارزیابی کمترین میزان خطا رادارند اگرچه در نتایج آزمون برای همین سهم در مرحله پنج با یکی از حالات پنج‌تایی پیش‌بینی با میزان خطای کم دیده نمی‌شود ولی در مرحله ۱ با تک ویژگی پیشنهادی شاخص نرخ تغییر معیارهای ارزیابی، کمترین میزان خطا را نشان می‌دهد

جدول ۳: نتایج ارزیابی برای شرکت نفت پارس با ویژگی‌های پیشنهادی

MSE اعتبار سنجی	RMSE اعتبار سنجی	الگوریتم	ویژگی
۰/۰۰۱۳۷	۰/۰۳۷۰۱	SVR	بازده کل،شاخص صنعت،بازده صنعت،میانگین متحرک۳۰روزه،بتاصنعت
۰/۰۰۱۵۴	۰/۰۳۹۲۴	MLP	شاخص کل، بازده کل، شاخص صنعت،شاخص توان نسبی
۰/۰۰۱۷۸	۰/۰۴۲۱۹	DT	بازده کل،شاخص صنعت،بازده صنعت،بتاصنعت
MSE آزمون	**RMSE آزمون**	**الگوریتم**	**ویژگی**
۰/۳۲۵۷	۰/۱۸۰۴۷	SVR	شاخص نرخ تغییر
۰/۰۵۱۷۲	۰/۲۲۷۴۲	MLP	آخرین روزهفته،بازده کل
۰/۳۹۳۲	۰/۱۹۸۳	DT	بتای صنعت،شاخص قیمت

درواقع برای نمونه‌های آزمون معیارهای ارزیابی در مرحله یک یا دو یعنی با تک ویژگی یا دو ویژگی کمترین میزان خطا را نشان می‌دهند. همان‌طور که در جدول (۴) و (۵) نشان داده‌شده است، برای سهام پتروشیمی شازند نتایج پیش‌بینی الگوریتم‌ها برای دادگان با ویژگی‌های پیشنهادی در مقایسه با دادگان اولیه با ده ویژگی پایه بهبود داشته است. برای این سهم نتایج اعتبارسنجی نشان می‌دهد الگوریتم SVR عملکرد بهتری نسبت به دو الگوریتم دیگر دارد.

جدول ۴: نتایج ارزیابی سهام پتروشیمی شازند با دادگان اولیه

MSE اعتبارسنجی	RMSE اعتبارسنجی	الگوریتم
۰/۰۰۵۸۰	۰/۰۷۶۱۵	SVR
۰/۰۰۴۷۸	۰/۰۶۹۱۳	MLP
۰/۰۰۵۹۹	۰/۰۷۷۳۹	DT
MSE آزمون	**RMSE آزمون**	**الگوریتم**
۰/۱۴۴۰۷	۰/۳۷۹۵۶	SVR
۰/۱۸۱۱۰	۰/۴۲۵۰۶	MLP
۰/۱۷۱۵۵	۰/۴۱۴۱۸	DT

جدول ۵: نتایج ارزیابی برای پتروشیمی شازند با ویژگی‌های پیشنهادی

MSE اعتبار سنجی	RMSE اعتبار سنجی	الگوریتم	ویژگی
۰/۰۰۱۰۱	۰/۰۳۱۷۸	SVR	شاخص کل، بازده کل، بازده صنعت، میانگین همگرایی واگرایی
۰/۰۰۱۸۸	۰/۰۴۳۳۵	MLP	شاخص صنعت، بازده صنعت، شاخص قیمت هموزن
۰/۰۰۲۰۴	۰/۰۴۵۱۶	DT	بازده کل، شاخص صنعت، بازده صنعت، بتای صنعت، میانگین متحرک۳روزه،باند بالابولینگر
MSE آزمون	**RMSE آزمون**	**الگوریتم**	**ویژگی**
۰/۱۰۶۱۲	۰/۳۲۵۷۷	SVR	بازده کل، شاخص نرخ تغییر
۰/۱۶۵۸۳	۰/۴۰۷۲۳	MLP	بازده کل،میانگین متحرک نمایی
۰/۱۶۶۵۸	۰/۴۰۸۱۴	DT	شاخص توان نسبی

www.SID.ir

Archive of SID

در آزمون هم مجدداً هر الگوریتم SVR کارایی بهتری از خود نسبت به درخت تصمیم و MLP نشان می‌دهد با این تفاوت که SVR در آزمون برای ویژگی‌های پیشنهادی بازده کل و شاخص نرخ تغییر (یکی از حالات دوتایی ویژگی‌ها در مرحله دو Forward) حداقل میزان خطا را دارد.

۶- نتیجه‌گیری

نتایج ارزیابی به روی دو سهم با روش انتخاب ویژگی پیشرو و پس‌رو برای ۲۳ ویژگی پیشنهادی که بر پایه تحلیل فنی می‌باشند چنین نشان دادند که کارایی الگوریتم‌های MLP و SVR نسبت به درخت تصمیم بالاتر می‌باشد. نتیجه‌ی قابل‌توجه دیگری که در ارزیابی‌ها مشاهده می‌شود این است که در اعتبارسنجی همه مراحل یک تا ۲۳ انتخاب ویژگی، خطای کمتری نسبت به دادگان اولیه دارد که با استفاده از انتخاب ویژگی پیشرو مراحل یک تا شش بهترین پیش‌بینی با حداقل میزان خطا ملاحظه می‌شود، از بین این مراحل، حالات چهارتایی یا پنج‌تایی یا شش‌تایی از بقیه‌ی حالات نتیجه‌ی بهتری داشتند. در مرحله آزمون اگرچه حداقل میزان خطا در حالات چهارتایی یا پنج‌تایی یا شش‌تایی از ویژگی-های پیشنهادی نبود، اما در مرحله یک و دو بودند، که این مراحل شامل مرحله یک تا شش نیز می‌باشد، درنتیجه نتایج آزمون قابل‌قبول است. ضمناً نتایج اعتبار سنجی به‌طور واضح نشان‌دهنده‌ی حضور مؤثر ویژگی‌های پیشنهادی شاخص کل، بازده کل، شاخص صنعت، بازده صنعت برای هر دو الگوریتم SVR و MLP در ترکیب‌های مختلف از ویژگی‌های انتخاب‌شده در هر دو سهم می‌باشد.

مراجع

[1] R. Gupta, C. Pierdzioch, R. Selmi, and M. E. Wohar, "Does partisan conflict predict a reduction in US stock market (realized) volatility? Evidence from a quantile-on-quantile regression model," *The North American Journal of Economics and Finance*, vol. 43, pp. 87–96, 2018.

[2] J. Nadkarni and R. Ferreira Neves, "Combining NeuroEvolution and Principal Component Analysis to trade in the financial markets," *Expert Systems with Applications*, vol. 103, pp. 184–195, 2018.

[3] K. Pilbeam, *Finance & financial markets*. Macmillan International Higher Education, 2018.

[4] X. Zhong and D. Enke, "Forecasting daily stock market return using dimensionality reduction," *Expert Systems with Applications*, vol. 67, pp. 126–139, 2017.

[5] M. Ballings, D. Van Den Poel, N. Hespeels, and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7046–7056, 2015.

[6] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 1, pp. 259–268, 2015.

[7] P. P.Edgar Torres, M. Hern, E. A. Torres, and S. G. Yoo, "Stock Market Data Prediction Using Machine Learning Techniques," vol. 1, pp. 539–547,2019.

[8] M. Ghiassi and C. Burnley, "Measuring effectiveness of a dynamic artificial neural network algorithm for classification problems," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3118–3128, 2010.

[9] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, "An innovative neural network approach for stock market prediction," *The Journal of Supercomputing*, 2018.

[10] M. W. Hsu, S. Lessmann, M. C. Sung, T. Ma, and J. E. V. Johnson, "Bridging the divide in financial market forecasting: machine learners vs. financial economists," *Expert Systems with Applications*, vol. 61, pp. 215–234, 2016.

[11] R. C. Cavalcante, R. C. Brasileiro, V. L. F. Souza, J. P. Nobrega, and A. L. I. Oliveira, "Computational Intelligence and Financial Markets: A Survey and Future Directions," *Expert Systems with Applications*, vol. 55, pp. 194–211, 2016.

[12] U. Gurav and N. Sidnal, "Predict Stock Market Behavior: Role of Machine Learning Algorithms," in *Intelligent Computing and Information and Communication*, Springer, 2018, pp. 383–394.

[13] Y. Chen and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Expert Systems with Applications*, vol. 80, pp. 340–355, 2017.

[14] S. Barak and M. Modarres, "Developing an approach to evaluate stocks by forecasting effective features with data mining methods," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1325–1339, 2015.

پاورقی‌ها:

[1] Multilayer Perceptron
[2] Decision tree
[3] Support vector machine

www.SID.ir