ISSN: 2645-4408
پژوهش در حسابداری و علوم اقتصادی
فصلنامه پژوهش در حسابداری و علوم اقتصادی
سال هشتم، شماره ۲ (پیاپی ۲۹)، بهار ۱۴۰۳

بررسی تاثیرات مدل رگرسیون خودهمبسته بر پیش‌بینی بازار سرمایه

سیدمحمد سجادی[1]*، جواد عین‌آبادی[2]

تاریخ دریافت: ۱۴۰۳/۰۳/۰۸
تاریخ پذیرش: ۱۴۰۳/۰۳/۲۵
کد مقاله: ۳۹۰۳۴

چکـیده

هدف: با توجه به اقبال عمومی به بازار سرمایه ایران و گسترش استفاده از سیستم‌های یادگیری ماشین در این پژوهش سعی بر بررسی کاربرد یکی از مهم‌ترین الگوریتم‌های یادگیری ماشین در بازار سرمایه ایران شد.
روش: داده‌های قیمت پایانی شاخص کل بورس تهران، شاخص هم‌وزن و شاخص ۳۰ شرکت بزرگ بازار در بازه زمانی بین سال ۱۳۹۵ تا سال ۱۴۰۲ در چارچوب زمانی روزانه جمع‌آوری شد. پس از تایید صحت اطلاعات، به کمک زبان برنامه‌نویسی پایتون و با استفاده از مدل خودهمبسته به بررسی عملکرد این الگوریتم روی این شاخص‌ها پرداخته شد. در هر مرحله بخشی از داده‌ها را به بخش یادگیری و بخشی دیگر را به ارزیابی و تست سیستم اختصاص داده شد.
یافته‌ها: نتایج بررسی نشان از کاربردی بودن مدل خودهمبسته در بازار سرمایه ایران دارد و به ترتیب شاخص کل و سپس شاخص هم‌وزن و در نهایت شاخص ۳۰ شرکت بزرگتر بهترین نتیجه را در این مدل داشته‌اند.
نتیجه‌گیری: نتایج پژوهش نشان داد که مدل خودهمبسته بهترین عملکرد را روی شاخص کل داشته و در کل استفاده از این مدل نمی‌تواند به تنهایی نتیجه بخش باشد ولی می‌تواند در کنار ابزارهای دیگر یک دید مناسب به سرمایه‌گذاران ارائه کند.

واژگـان کلـیدی: پیش‌بینی بازار سرمایه، بورس اوراق بهادار، سرمایه‌گذاری، مدل خودهمسبته

۱- دانشجوی کارشناسی ارشد، دانشکده حسابداری و مالی، موسسه آموزش عالی الکترونیکی ایرانیان، تهران، ایران (نویسنده مسئول)
mohammadsjh@gmail.com
۲- استادیار، دانشکده حسابداری و مالی، موسسه آموزش عالی الکترونیکی ایرانیان، تهران، ایران

۱- مقدمه

بازار سرمایه یکی از مهم‌ترین بازارهای مالی هر کشور و اقتصاد است که رشد اقتصادی هر کشور نیازمند بازار سرمایه‌ای کارا و کارآمد است که منابع و دارایی‌ها به صورت بهینه به بخش‌های مناسب هر صنعت هدایت شود. بازار سهام به عنوان یک سیستم پیچیده و پویا، تحت تأثیر عوامل متعددی از جمله رویدادهای اقتصادی و سیاسی، رفتار سرمایه‌گذاران و عوامل فنی قرار می‌گیرد. پیش‌بینی درست و دقیق رفتار بازار سهام، علمی بسیار ارزشمند است و به سرمایه‌گذاران، مدیران صندوق‌های سرمایه‌گذاری، تحلیلگران مالی و سایر علاقه‌مندان در اتخاذ تصمیمات سرمایه‌گذاری و مدیریت ریسک کمک قابل توجهی می‌کند. در حال حاضر ارزش دلاری بورس تهران در حدود ۱۵۰ میلیارد دلار است و تا به امروز بالاترین ارزش دلاری بورس تهران ۴۲۷ میلیارد دلار بوده است که در نیمه اول سال ۱۳۹۹ ثبت شده است. ارزش معاملات خرد سهام و صندوق‌ها به صورت میانگین بین ۳ همت(هزار میلیارد تومان) تا ۱۰ همت برآورد می‌شود.

پیشبینی بازار سهام و تاثیرپذیری این بازار از عوامل مختلف بارها در مقالات مختلف مورد بحث و بررسی قرار گرفته است. نسبت‌های مرتبط با سنجش فعالیت شرکت و نسبت‌های مرتبط با اندازه‌گیری وضعیت بدهی‌ها، اندازه شرکت و نوع صنعت، در پیشبینی رتبه عملکرد شرکت‌ها مفید بوده‌اند (بهرامفر و ساعی، ۱۳۸۵). عوامل مختلفی در پیش‌بینی قیمت‌های آتی در بازار سهام توسط الگوریتم‌های پیش‌بینی‌کنند دخیل هستند. میان قیمت های گذشته سهام, درصد سهام مبادله شده, شاخص بازده نقدی و قیمت و تغییرات آتی قیمت سهام رابطه معنی داری وجود دارد (معصوم‌زاده و قالیباف اصل، ۱۳۸۹). تفاوت زیادی بین تحلیل بازار سهام ایران و بازار فارکس وجود دارد. به طوری که روش تحلیل بازار بورس ایران بیشتر به مسائل داخلی کشور ایران و یک صنعت و یا یک شرکت خاص بر می‌گردد و نوع سرمایه گذاری در آن بلندمدت است. ولی تحلیل بازار فارکس به مسائل کلی جهان و شاخص های کلان اقتصادی جهان و بیشتر به روانشناسی و مدیریت سرمایه برمی‌گردد و نوع سرمایه‌گذاری در آن کوتاه مدت است و در کل فلسفه بازار بورس اوراق بهادار تهران سرمایه‌گذاری است ولی فلسفه بازار فارکس نوسانگیری است (فرهنگ دوست، ۱۴۰۱).

به دلیل حجم بالای اطلاعات و تغییرات روزانه شاخص‌ها، استفاده از الگوریتم‌های پیشبینی یک امر ضروری تلقی می‌شود. در این تحقیق سعی شده است به بررسی مواردی اعم از، بررسی فرضیه پیشبینی‌پذیری بازار سهام و بررسی نتایج استفاده از الگوریتم‌های هوش مصنوعی در بازار سهام بپردازیم. سری زمانی یک مجموعه مرتب از داده‌ها است که بر اساس زمان به ترتیب قرار دارند. پیشبینی قیمت‌ها با استفاده از سری زمانی، یعنی پیش‌بینی مقادیر آینده در یک دنباله زمانی. این موضوع در حوزه تحلیل فنی و مالی بسیار مورد توجه است. در پیشبینی قیمت‌ها با استفاده از سری زمانی، معمولاً از مدل‌های آماری و ریاضی برای تحلیل و پیش‌بینی استفاده می‌شود. این مدل‌ها می‌توانند شامل مدل‌های خطی مانند مدل ARIMA (مدل خودرگرسیوی متحرک متوالی) و مدل‌های غیرخطی مانند مدل‌های GARCH (مدل همتراکمی تجزیه شده خودرگرسیوی عمومی) باشند. با استفاده از سری زمانی، می‌توان الگوها، تغییرات و روند قیمت‌ها را شناسایی کرده و براساس آن‌ها به پیش‌بینی قیمت‌های آینده بپردازیم. عوامل متعددی می‌توانند تحت تأثیر قرار گیرند، از جمله تغییرات اقتصادی، سیاسی، اطلاعات عمومی، رفتار سرمایه‌گذاران و الگوهای فصلی. از تحلیل و پیش‌بینی سری زمانی می‌توان در تصمیم‌گیری‌های سرمایه‌گذاری، مدیریت ریسک و بهره‌وری بیشتر استفاده کرد.

رگرسیون[1]، یک روش آماری است که برای بررسی رابطه بین یک متغیر وابسته(متغیر پاسخ) و یک یا چند متغیر مستقل(متغیرهای توضیحی) استفاده می‌شود. در رگرسیون خطی، فرض می‌شود که رابطه بین متغیرهای وابسته و مستقل به صورت خطی است. استفاده از رگرسیون ابتدا از محاسبات و مقاله (فرانسیس گالتون[2]، ۱۸۸۶) شروع شد. (هوک، ۲۰۰۹) بیان کرد که بعد از استفاده‌های بسیار از یادگیری ماشین در حوزه‌هایی مانند پردازش زبان طبیعی، احتمال داده شد که از این روش‌ها در فهمیدن بازارهای مالی نیز استفاده کرد. در برخی مواقع، رابطه بین متغیرهای وابسته و مستقل به صورت غیرخطی یا تعاملی است. در این حالت، سیستم‌های رگرسیون خود همبسته[3] مورد استفاده قرار می‌گیرند. در این سیستم‌ها، متغیرهای وابسته در زمان‌های گذشته تأثیری بر روند تغییرات خود در زمان فعلی دارند.

۲- مبانی نظری و پیشینه پژوهش

پیشبینی بازارهای مالی امری جذاب و در نهایت اگر منجر به پیدایش یک سبک مناسب معاملاتی شود راهکاری مناسب برای کسب سود در بازارهای مالی است. از این رو سیستم‌های متعددی برای بررسی و پیشبینی بازارهای مالی در طی سالیان گذشته

1 Regression
2 Francis Galton
3 autoregressive models

توسعه داده شده است. بعضی از این سیستم‌های گذشته‌نگر[1] و بعضی از آن‌ها آینده‌نگر[2] هستند. سیستم‌های یادگیری ماشین[3] با توجه به دریافت اطلاعات گذشته بازار سعی در کشف آینده قیمتی دارند.

(دنگ و ساکورایی[4]، ۲۰۱۳) در مقاله "قوانین بازار مبادله ارزهای خارجی با استفاده از یک اندیکاتور و چند چارچوب زمانی" از الگوریتم ژنتیک و اندیکاتور RSI برای ایجاد قوانین معاملاتی در چارچوب زمانی ۱ ساعته استفاده کردند و روی جفت ارز یورو/دلار در بازه سال ۲۰۱۱ در بهترین حالت به بازدهی بیشتر از ۱۳۸۰ پیپ[5] دست یافتند. (تهرانی و همکاران، ۱۳۸۹) طی مطالعات‌شان برای پیش‌بینی نوسان شاخص بازده نقدی و قیمت بورس تهران(TEDPIX) بر اساس معیار ارزیابی میانگین مربعات خطا(RMSE) به این نتیجه رسیدند که عملکرد مدل میانگین متحرک ۲۵۰ روزه، هموارسازی نمایی و CGARCH طبق معیار RMSE از دیگر مدل‌ها بهتر است. (غفاری و یوسفی، ۱۳۹۰) در مقاله "پیشبینی نرخ ارز با استفاده از شبکه عصبی" به این نتیجه رسیدند که استفاده از شبکه‌های عصبی قدرت پیشبینی مناسبی را از نرخ ارزها دارد. (حیدری و امیری، ۱۴۰۱) به بررسی قدرت مدل‌های مبتنی بر هوش مصنوعی در پیشبینی روند قیمت سهام بورس اوراق بهادار پرداختند و نتیجه گرفتند که مدل‌های مبتنی بر یادگیری عمیق نسبت به سایر مدل‌ها عملکرد بهتری از خود نشان می‌دهند و در پیش‌بینی روند کوتاه‌مدت قیمت سهام، از دقتی حدود ۷۰ تا ۸۰ درصد برخوردارند. (فشاری و مظاهری‌فر، ۱۳۹۵) بیان کردند که شبکه‌های عصبی توانسته عملکرد بهتری را در پیشبینی بازده اوراق بهادار نسبت به سیستم فازی عصبی نشان دهد و همچنین در بررسی عملکرد سه الگوریتم کوادراتیک، ژنتیک و ممتیک، نتایج نشان می‌دهد که الگوریتم جهش ترکیبی قورباغه توانسته عملکرد و نتیجه بهتری را در مقایسه با الگوریتم ژنتیک نسبت به الگوریتم کوادراتیک نشان دهد. (شاه‌حسینی و رضایی، ۱۳۹۶) در مقاله "پیشبینی نرخ رسمی ارز در ایران همراه با عاملهای ARIMA با استفاده از مدل خودرگرسیونی مداخلهای و مقایسهی آن با مدل گام تصادفی" با مقایسه مدل خود رگرسیونی ARIMA با مدل گام‌برداری تصادفی برای مدل سازی نرخ برابری ریال به دلار برای سال های ۱۳۵۷ تا ۱۳۹۴ به این نتیجه رسیدند که مدل خودرگرسیونی به نسبت مدل گام‌برداری تصادفی برتری دارد. (عرفانی، ۱۳۸۸) نشان داد که قدرت پیشبینی مدل ARFIMA در قیاس با مدل ARIMA بر شاخص کل بورس اوراق بهادار تهران بیشتر است.

در این پژوهش به بررسی سیستم خودهمبسته برای بررسی بازار سرمایه ایران روی چند شاخص کلیدی بازار می‌پردازیم. برای رسیدن به مقصود نهایی ابتدا به روش‌شناسی و بیان فرضیه می‌پردازیم و سپس به معرفی داده‌های استفاده شده و آزمون‌هایی که روی این داده‌ها انجام شده است می‌پردازیم . در انتها نتیجه‌گیری و پیشنهاداتی ارائه خواهد شد.

۳- روش‌شناسی

در این پژوهش به بررسی نتایج حاصل از تخمین قیمت پایانی روز بعد به وسیله الگوریتم رگرسیون اتوماتیک پرداختیم که در ادامه به تفضیل بررسی خواهد شد.

- در گام اول تمام داده‌های عدد پایانی در شاخص‌های مورد بررسی دریافت می‌شود.
- سپس نوبت به پیاده‌سازی تابع سری زمانی و تابع لگ سری زمانی می‌رسد.
- سپس داده‌های هر شاخص به دو دسته تقسیم می‌شوند. عملیات تفکیک یا تقسیم داده‌ها در یادگیری ماشین روشی برای سنجش کیفیت عملکرد یک الگوریتم یادگیری ماشین است. در این پژوهش ۸۰ درصد از داده‌ها برای آموزش و ۲۰ درصد از داده‌ها برای تست استفاده شده است.
- در قدم بعدی تابع خودهمبسته پیاده‌سازی می‌شود.
- در مرحله آخر معیارهای مورد نیاز برای برآورد خطای مدل پیاده‌سازی می‌شوند که نتایج با مدل‌های دیگر و با بازه‌های زمانی دیگر مقایسه شوند.

قدم‌های فوق در ادامه به تفصل و تک تک با بررسی مفاهیم و فرمول‌ها بررسی می‌شوند.

۳-۱- فرضیه پژوهش

الگوریتم رگسیون خودهمبسته توانایی پیشبینی شاخص‌های بازار سهام ایران را دارد.

1 Retrospective systems
2 Prospective systems
3 Machine Learning
4 Dang & Sakurai
5 pip

۳-۲- رگرسیون

رگرسیون یکی از روش‌های تحلیل آماری است که برای بررسی و پیش‌بینی رابطه بین یک متغیر وابسته(متغیر پاسخ) و یک یا چند متغیر مستقل(متغیرهای توضیحی) استفاده می‌شود. رگرسیون به ما اجازه می‌دهد تا با استفاده از داده‌های موجود، یک مدل ریاضی ایجاد کنیم که بتوانیم با استفاده از آن، مقدار متغیر پاسخ را براساس مقادیر متغیرهای توضیحی پیش‌بینی کنیم. برای مثال برای اگر بخواهیم تاثیر حجم را بر قیمت مشخص کنیم، حجم یک متغیر توضیحی است و قیمت یک متغیر وابسته است.
اگر برای شناسایی و پیش‌بینی متغیر وابسته فقط از یک متغیر مستقل استفاده شود، این مدل را رگرسیون خطی ساده می‌گوییم. فرم مدل رگرسیون خطی ساده به شکل زیر است:

$$Y = \beta_1+\beta_1 X + \epsilon \quad (۱)$$

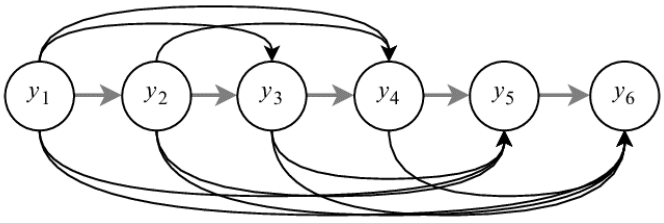
این رابطه، معادله یک خط است که جمله خطا یا همان (ϵ) به آن اضافه شده. پارامترهای این مدل خطی عرض از مبدا (β۰) و شیب خط (β۱) است. شیب خط در حالت رگرسیون خطی ساده، نشان می‌دهد که میزان حساسیت متغیر وابسته به متغیر مستقل چقدر است. به این معنی که با افزایش یک واحد به مقدار متغیر مستقل چه میزان متغیر وابسته تغییر خواهد کرد. عرض از مبدا نیز بیانگر مقداری از متغیر وابسته است که به ازاء مقدار متغیر مستقل برابر با صفر محاسبه می‌شود. به شکل دیگر می‌توان مقدار ثابت یا عرض از مبدا را مقدار متوسط متغیر وابسته به ازاء حذف متغیر مستقل در نظر گرفت.

۳-۳- سری‌های زمانی

سری زمانی مجموعه‌ای از داده‌ها است که به ترتیب زمانی ثبت شده‌اند. در سری زمانی، هر داده مربوط به یک زمان مشخص است و معمولاً به صورت متوالی ثبت می‌شود. مثال‌هایی از سری‌های زمانی شامل داده‌های روزانه قیمت سهام، دماهای روزانه، ترافیک روزانه و تعداد فروش روزانه می‌باشد. در سری‌های زمانی، زمان معمولاً به عنوان محور افقی نمودارها و نمایش‌های دیگر استفاده می‌شود. با تحلیل سری‌های زمانی، می‌توان الگوها، روندها، فصلیت‌ها و تغییرات در طول زمان را شناسایی کرد. این تحلیل می‌تواند در پیش‌بینی رویدادها، تصمیم‌گیری‌های کسب و کار، برنامه‌ریزی منابع و مدیریت ریسک مورد استفاده قرار بگیرد. برای تحلیل سری‌های زمانی، روش‌های مختلفی وجود دارد از جمله تجزیه‌وتحلیل آماری، روش‌های همبستگی، مدل‌های زمانی، شبکه‌های عصبی و الگوریتم‌های یادگیری ماشین. این روش‌ها بسته به مسئله مورد بررسی و خصوصیات سری زمانی مورد استفاده قرار می‌گیرند. اگر بین ویژگی‌های مختلف یک پدیده فقط از یکی ویژگی برای ایجاد مدل سری زمانی استفاده شود، مدل را یک متغیره می‌نامند. ولی اگر از چندین ویژگی برای ایجاد مدل سری زمانی استفاده شود، مدل سری زمانی را چند متغیره می‌گویند.

۳-۴- مدل خودهمبسته

این مدل یک روش پیشبینی در سری‌های زمانی است. مدل‌های خودهمبسته، دسته‌ای از مدل‌های آماری هستند که بر اساس مقادیر قبلی خود، برای پیش‌بینی یا تولید دنباله‌های نقاط داده‌ای استفاده می‌کنند. این مدل‌ها فرض می‌کنند که هر نقطه داده در یک دنباله به صورت خطی به تعداد ثابتی از نقاط داده قبلی وابسته است. اصطلاح "خودهمبسته" به این موضوع اشاره دارد که مدل بر مقادیر گذشته همان متغیر، رگرسیون می‌کند.



شکل ۱ - مدل مفهومی خودهمبسته

از آنجایی که این مدل به مانند یک مدل رگرسیون نوشته شده، به آن مدل خودهمبسته(اتورگرسیو) می‌گویند، با توجه به اینکه برای $X(t)$ رگرسیون روی مقدارهای گذشته ایجاد شده، شرط استقلال متغیرهای توصیفی حداقل برای تعداد دسته‌های کوچکتر از p وجود نخواهد داشت. همچنین مقدار حال حاضر سری زمانی فقط به p مقدار قبلی وابسته بوده و به قبل از آن ارتباطی ندارد.
در زمینه مدل‌سازی زبان، مدل‌های خودهمبسته به طور رایج برای تولید متن با پیش‌بینی کلمه یا حرف بعدی در یک دنباله بر اساس کلمات یا حروف قبلی استفاده می‌شوند. آن‌ها با استفاده از مجموعه‌های بزرگی از داده‌های متنی آموزش می‌بینند.

اگر سری X را به شکل $x_1, x_2, \ldots, x_{n-1}, x_n$ داشته باشیم، در این صورت یک مدل خودهمبسته از مرتبه P به شکل زیر خواهد بود:

$$X_t = c + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \epsilon_t = c + \sum_{i=1}^{p} \varphi_i x_{t-i} + \epsilon_t \quad (۲)$$

در این رابطه، بردار ϕ ضرایب مربوط به نقاط قبلی بوده و c عددی ثابت است. مقدار ε_t نیز نشان‌دهنده اندک خطای مدل از مقادیر مشاهده شده است.

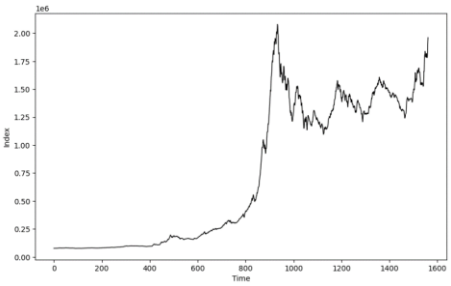
۳-۵- آموزش داده‌ها و تست

در یادگیری ماشین، بخش بندی داده‌ها به داده‌های آموزش و تست یک مرحله کلیدی است. این بخش بندی به ما امکان می‌دهد که مدل‌های یادگیری ماشین را با استفاده از داده‌های آموزش آموزش دهیم و سپس با استفاده از داده‌های تست عملکرد آن‌ها را ارزیابی کنیم. بنابراین، بخش بندی داده‌ها مهم است تا بتوانیم اعتبارسنجی مناسبی برای مدل‌های یادگیری ماشین‌مان انجام دهیم و از افتراق بین عملکرد مدل در داده‌های آموزش و تست مطمئن شویم.

معمولاً، داده‌ها را به صورت تصادفی به دو بخش آموزش و تست تقسیم می‌کنیم. بخش آموزش شامل داده‌هایی است که برای آموزش مدل‌های یادگیری ماشین استفاده می‌شوند. این داده‌ها به مدل کمک می‌کنند تا الگوها و قواعد موجود در داده‌ها را بیاموزد. بخش تست شامل داده‌هایی است که برای ارزیابی و اعتبارسنجی مدل‌های یادگیری ماشین استفاده می‌شوند. این داده‌ها برای اندازه‌گیری دقت و عملکرد مدل در پیش‌بینی داده‌های جدید استفاده می‌شوند.

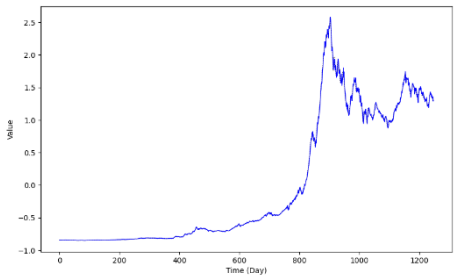
هنگام بخش بندی داده‌ها، باید توجه داشت که نمونه‌ها به صورت تصادفی و بدون تمایز در دو بخش آموزش و تست تقسیم شوند. همچنین، بهتر است حجم داده‌های آموزش بیشتر از داده‌های تست باشد تا مدل‌ها به طور کامل آموزش ببینند و از بیش‌برازش[1] جلوگیری شود.

ما در این پژوهش از بخش داده‌های اول، ۸۰ درصد داده‌ها را به بخش آموزش و ۲۰ درصد از داده‌ها را برای تست اختصاص دادیم. در تصویر ۲ نمودار کامل شاخص‌کل بورس را در بازه ۰۱-۰۷-۱۳۹۵ تا ۰۱-۰۱-۱۴۰۲ مشاهده می‌کنید.

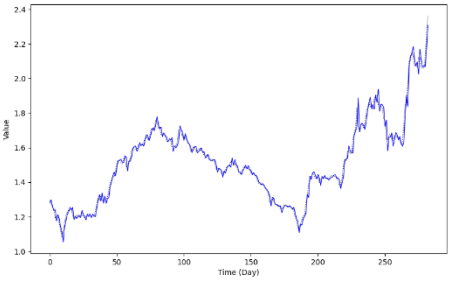


شکل ۲ - نمودار شاخص کل بورس در بازه اول مهر ۱۳۹۵ تا اول فروردین ۱۴۰۲

در نمودار زیر بخش آموزش داده‌های شاخص را مشاهده می‌کنید که ۸۰ درصد کل داده‌ها است:



شکل ۳ - نمودار داده‌های بخش آموزش در شاخص کل



شکل ٤ - نمودار داده‌های تست در شاخص کل

۳-۶- معیارهای ارزیابی

برای بررسی نتایج و ویژگی‌های مورد انتظار از معیارهای پرکاربرد و مطرح زیر که در ارزیابی مسائل رگرسیون بسیار پرکاربرد هستند استفاده می‌کنیم. هدف استفاده از این معیارها مقایسه مدل‌ها با یکدیگر و پیشبینی خطاهای مدل در مواجه با داده‌های جدید است.

۳-۶-۱- میانگین مربعات خطا (MSE)[2]

این معیار دارای بُعد است، به همین دلیل به عنوان گزارش نهایی در مسائل رگرسیون مناسب نیست.

1 overfitting
2 Mean Squared Error

$$MSE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \quad (۳)$$

۳-۶-۲- جذر میانگین مربعات خطا (RMSE)[1]

مزیت استفاده از این معیار یکسان بودن بعد و مقیاس آن با ویژگی هدف است. زیرا با جذر گرفتن از میانگین مربعات خطا به یک بعد یکسان با ویژگی هدف می‌رسیم.

$$RMSE(y,\hat{y}) = \sqrt{MSE(y-\hat{y})} \quad (۴)$$

۳-۶-۳- جذر مربعات خطای نرمال‌شده (NRMSE)[2]

این معیار، حاصل تقسیم جذر میانگین مربعات خطا بر بازه ویژگی هدف است، به همین دلیل عددی بدون بُعد و واحد بوده و برای گزارش بسیار مناسب است.

$$NRMSE(y,\hat{y}) = \frac{RMSE(y-\hat{y})}{Max(y)-Min(y)} \quad (۵)$$

۳-۶-۴- میانگین قدرمطلق خطا (MAE)[3]

این معیار به جای به توان ۲ رساندن خطاها، از تابع قدرمطلق استفاده می‌کند. بعد این معیار با ویژگی هدف یکسان است.

$$MAE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i| \quad (۶)$$

۳-۶-۵- میانگین درصد قدرمطلق خطا (MAPE)[4]

این معیار مشابه میانگین قدرمطلق خطا است، اما به جای خطا، از خطای نسبی استفاده شده است و معیاری بدون واحد است.

$$MAPE(y,\hat{y}) = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{e_i}{y_i}\right| \quad (۷)$$

۳-۶-۶- ضریب تعیین یا امتیاز (R^2)[5]

این معیار همواره عددی کوچک‌تر از ۱ است. اگر مدلی همواره میانگین ویژگی هدف را در خروجی تولید کند، مقدار این معیار برابر صفر خواهد بود. ضریب تعیین معمولاً به صورت درصد بیان می‌شود. از ضریب تعیین تنها برای مقایسه مدل‌ها و گزارش نتایج استفاده می‌شود.

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1 - \frac{MSE(y,\hat{y})}{\sigma^2(y)} \quad (۸)$$

۴- یافته‌های پژوهش

با توجه به مدل و الگوریتم شرح داده شده اطلاعات عدد پایانی شاخص کل ، شاخص کل هم‌وزن و شاخص ۳۰ شرکت بزرگ در بازه زمانی اول مهر ۱۳۹۵ تا اول فروردین ۱۴۰۲ به عنوان ورودی به مدل داده شد. (نمونه‌ای از این اطلاعات را که مربوط به شاخص کل هم‌وزن است را در جدول زیر مشاهده می‌کنید.)

1 Root Mean Squared Error
2 Normalized Root Mean Squared Error
3 Mean Absolute Error
4 Mean Absolute Percentage Error
5 Coefficient of Determination

	Date	Weekday	Open	High	Low	Close		Adj Close	Volume
J-Date							J-Date		
1398-01-05	2019-03-25	Monday	33024.0	33391.0	33024.0	33391.0	1398-01-05	33390.9	1030038293
1398-01-06	2019-03-26	Tuesday	33569.0	34043.0	33569.0	34034.0	1398-01-06	34034.1	1509089603
1398-01-07	2019-03-27	Wednesday	34226.0	34406.0	34226.0	34406.0	1398-01-07	34406.1	1807050103
1398-01-10	2019-03-30	Saturday	34552.0	35084.0	34552.0	35084.0	1398-01-10	35084.1	1856911671
1398-01-11	2019-03-31	Sunday	35288.0	35562.0	35288.0	35368.0	1398-01-11	35367.9	1926218389
...	...	...	...	...	...	...	...	...	...
1398-12-24	2020-03-14	Saturday	179275.0	179275.0	175725.0	175731.0	1398-12-24	175731.0	4006156865
1398-12-25	2020-03-15	Sunday	174963.0	174963.0	172385.0	172406.0	1398-12-25	172406.0	4765050083
1398-12-26	2020-03-16	Monday	172501.0	174668.0	172501.0	174668.0	1398-12-26	174668.0	4543347371
1398-12-27	2020-03-17	Tuesday	174967.0	175025.0	173504.0	173784.0	1398-12-27	173784.0	3263499294
1398-12-28	2020-03-18	Wednesday	174329.0	177084.0	174329.0	177080.0	1398-12-28	177080.0	4058241862

شکل ۵ – نمونه اطلاعات ورودی مدل

جدول ۱ – آماره‌های داده‌های هر شاخص

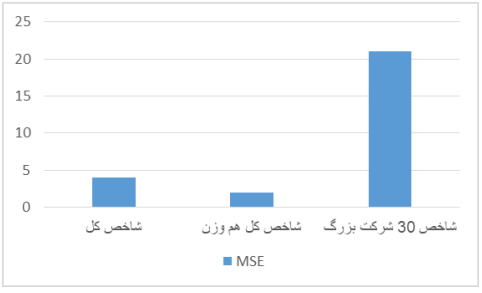
شاخص	واریانس	انحراف معیار
شاخص کل	402739413798.3	634617.5
شاخص هم‌وزن	36018631918.9	189785.7
شاخص ۳۰ شرکت بزرگ	۱۲۴۲۶۲۳۸۲۴,۹	35250.8

جدول ۲ – نتایج معیارها در مجموعه داده‌های آموزش

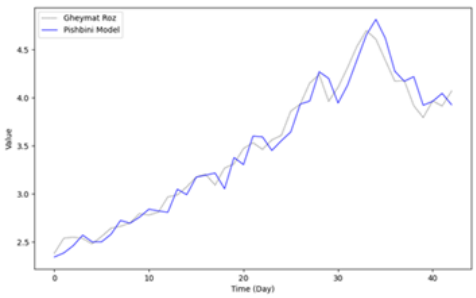
شاخص	MSE	RMSE	NRMSE	MAE	MAPE	R^2
شاخص کل	0.0004	0.0209	0.61 %	0.0111	2.28 %	99.96 %
شاخص هم‌وزن	0.0002	0.0144	0.5 %	0.0078	1.68 %	99.98 %
شاخص ۳۰ شرکت بزرگ	0.0021	0.0453	1.33 %	0.0177	2.2 %	99.79 %

جدول ۳ – نتایج معیارها در مجموعه داده‌های تست

شاخص	MSE	RMSE	NRMSE	MAE	MAPE	R2
شاخص کل	0.001	0.0312	2.43 %	0.0214	1.38 %	98.29 %
شاخص هم‌وزن	0.0007	0.0265	1.77 %	0.0185	1.32 %	99.32 %
شاخص ۳۰ شرکت بزرگ	0.0022	0.0466	2.65 %	0.0313	1.7 %	97.89 %



شکل ٦ – مقایسه شاخص MSE (ضرب در ۱۰۰۰۰ برای مقیاس بهتر در نمودار)



شکل ۷ – نمونه پیشبینی مجموعه داده تست در شاخص کل

۵- نتیجه‌گیری و پیشنهادها

نتایج به‌دست آمده از تحلیل اطلاعات ورودی به مدل به شرح زیر است:

- نتایج فاکتورهای به‌دست آمده در هر دو مجموعه داده آموزش و تست مناسب است و اختلاف عجیبی بین داده‌های آموزش و داده‌های تست وجود ندارد.

- نتایج به‌دست آمده از معیارهای MSE و R^2 نشان دهنده عملکرد مناسب این مدل دارد که در قیاس با مدل‌ها و الگوریتم‌های دیگر قابل دفاع است.
- در اکثر معیارها شاخص کل به نسبت دو شاخص دیگر عملکرد بهتری داشته است.
- با کاهش میزان مجموعه داده‌های تست، معیارهای هر شاخص عملکرد بدتری را ثبت می‌کند.
- با افزایش میزان داده‌های اولیه نتایج، معیارها در تمامی شاخص‌ها نتایج بسیار بهتری را ثبت می‌کنند.
- به دلیل رشد بسیار شدید شاخص‌ها در سال‌ها ابتدایی داده‌های کوچکتر تقریبا همگی در مجموعه داده‌های آموزش قرار دارند.
- نتایج مدل نشان از این دارد که این مدل می‌تواند در کنار سایر روش‌های تحلیلی اعم روش تحلیل بنیادی، تحلیل تکنیکال، تحلیل رفتاری و ... که سرمایه‌گذاران برای خرید و فروش در بازارهای مالی از آن‌ها بهره می‌گیرند، مورد استفاده قرار بگیرد و این مدل صرفا یک بخش از تحلیل و یاری‌رسان سرمایه‌گذاران در بازار سرمایه ایران باشد.

در نهایت پیشنهاد می‌کنیم برای بررسی‌های بیشتر از ترکیب مدل میانگین متحرک و مدل خودهمبسته استفاده کنید.

منابع

۱. بهرام‌فر تقی، ساعی محمدجواد، (۱۳۸۵)، «ارایه مدل برای پیشبینی عملکرد (مالی و بازار) شرکتهای پذیرفته شده در بورس اوراق بهادار تهران با استفاده از اطلاعات مالی منتشره»، بررسی‌های حسابداری و حسابرسی، دوره ۱۳، شماره ۴۳، صص ۷۰-۴۵

۲. تهرانی رضا، محمدی شاپور، پورابراهیمی محمدرضا، (۱۳۸۹)، «مدل‌سازی و پیش‌بینی نوسانات بازده در بورس اوراق بهادار تهران»، تحقیقات مالی، دوره ۱۲، شماره ۳۰، صص ۲۳-۳۶

۳. حیدری مهدی، امیری حمیدرضا، (۱۴۰۱)، «بررسی قدرت مدل‌های مبتنی بر هوش مصنوعی در پیش‌بینی روند قیمت سهام بورس اوراق بهادار تهران»، تحقیقات مالی، دوره ۲۴، شماره ۴، صص ۶۰۲-۶۲۳

۴. شاه‌حسینی سمیه، رضایی علی، (۱۳۹۶)، «پیشبینی نرخ رسمی ارز در ایران همراه با عاملهای ARIMA با استفاده از مدل خودرگرسیونی مداخلهای و مقایسهی آن با مدل گام تصادفی»، اقتصاد و تجارت نوین(پژوهشگاه علوم انسانی و مطالعات فرهنگی) ، دوره ۱۳، شماره ۴، صص۴۱-۵۰

۵. عرفانی علیرضا، (۱۳۸۸)، «پیش‌بینی شاخص کل بورس اوراق بهادار تهران با مدل ARFIMA»، فصلنامه تحقیقات اقتصادی، دوره ۴۴، شماره ۱

۶. غفاری مهدی، یوسفی راحله، (۱۳۹۰)، «مدلسازی پیش بینی قیمت ارز با استفاده از شبکه های عصبی»، مهندسی مالی و مدیریت اوراق بهادار (مدیریت پرتفوی)، دوره ۲، شماره ۸، صص ۹۹-۱۲۰

۷. فشاری مجید، مظاهری‌فر پوریا، (۱۳۸۵)، «مقایسه الگوریتمهای پیش بینی و بهینه سازی در بورس اوراق بهادار تهران»، فصلنامهٔ سیاستگذاری پیشرفت اقتصادی دانشگاه الزهرا، دوره ۴، شماره ۱۱

۸. فرهنگ‌دوست محمد، (۱۴۰۱)، « مقایسه روش تحلیل بازار بورس اوراق بهادار و بازار فارکس»، فصلنامه رویکردهای پژوهشی نوین در مدیریت و حسابداری، دوره ۶ شماره ۸۴، صص ۳۲۱-۳۲۸

۹. معصوم زاده نسیم، قالیباف اصل سیدحسن، (۱۳۸۹)، «پیش بینی احتمال تغییر قیمت سهام با استفاده از رگرسیون لجستیک در بورس اوراق بهادار تهران»، تحقیقات حسابداری و حسابرسی، دوره ۲، شماره ۵، صص ۵۳-۷۱

10. Deng, S., and Sakurai, A. (2013). Foreign Exchange Trading Rules Using a Single Technical Indicator from Multiple Timeframes. 2013 27th International Conference on Advanced Information Networking and Applications Workshops.
11. Galton, F. (1886). Regression Towards Mediocrity in Hereditary Stature. The Journal of the Anthropological Institute of Great Britain and Ireland 15 .pp. 246–263
12. Huck, N. (2009). Pairs selection and outranking: An application to the S&P 100 index. European Journal of Operational Research 196.pp. 819-825