Shahid Beheshti University
Journal of Financial Management Perspective
No. 34, Summer 2021
PP: 35-67

P-ISSN: 2645-4637
E-ISSN: 2645-4645

The Proposed Algorithm to Select Appropriate Features for Predicting Tehran Stock Exchange Index

Somayeh Mohebi*, Mohamad Esmaeil Fadaeinejad **, Mohammad reza Hamidizadeh***

Research Paper

Abstract

The performance of an intelligent model largely depends on the selection of the most relevant and most influential input variables and the lowest complexity of the learning model. Therefore, in the present study, to predict the index of Tehran Stock Exchange based on financial and economic variables, first prioritize features With MID algorithm, then 4 different neural network models (MLP, SVR, RBF, DNN) are used, which are the most important and innovative prediction models. According to the results of the analysis of the studied models, an algorithm is proposed to select the appropriate features on the index, as ISF-MID, and are compared with several similar methods. The data used in this study were collected daily in the period of 18/01/2014 to 21/08/2018. Evaluation of the models was performed by K-fold cross validation method. The MAE, MSE, and RMSE criteria are also used to evaluate the performance of the mentioned models. The results show that with the proposed method, with 7 selected features, it is possible to achieve high accuracy in predicting the daily index of the Tehran Stock Exchange.

Keywords: Stock Index Prediction; Feature Selection Algorithm; Radial Basis Function; Support Vector Regression; Deep Neural Network.

Received: 2021.February.21, Accepted: 2021. August.31

* Ph.D. Candidate in Financial Management, Shahid Beheshti University, Tehran, Iran.
** Associate Prof., Department of Financial Management and Insurance, Shahid Beheshti University, Tehran, Iran. (Corresponding Author). E-Mail: m-fadaei@sbu.ac.ir
*** Professor, Department of Business Management, Shahid Beheshti University, Tehran, Iran.

دانشگاه شهیدبهشتی
نشریه چشم‌انداز مدیریت مالی
شماره ۳۴ ـ تابستان ۱۴۰۰
ص‌ص ۶۷-۳۵

شاپای چاپی: ۲۶۴۵-۴۶۳۷
شاپای الکترونیکی: ۲۶۴۵-۴۶۴۵

الگوریتم پیشنهادی برای انتخاب ویژگی‌های مناسب به‌منظور پیش‌بینی شاخص بورس اوراق بهادار تهران

سمیه محبی*، محمداسماعیل فدائی‌نژاد، محمدرضا حمیدی‌زاده*****

مقاله پژوهشی

چکیده

عملکرد یک مدل هوشـــمند تا حد زیادی به انتخاب مرتبط‌ترین و تأثیرگذارترین متغیرهای ورودی و کمترین پیچیدگی مدل یادگیری بستگی دارد. از این‌رو در مطالعه حاضر، برای پیش‌بینی روزانه شاخص کل بورس اوراق بهادار تهران بر اساس متغیرهای مالی و اقتصادی، ابتدا اقدام به اولویت‌بندی ویژگی‌ها با الگوریتم MID نموده، سپس از ۴ مدل مختلف شبکه عصبی (MLP, SVR, RBF, DNN) که از مهم‌ترین و بدیع‌ترین مدل‌های پیش‌بینی می‌باشند، استفاده می‌شود. با توجه به نتایج بدست آمده از تحلیل مدل‌های مورد بررســـی، در نهایت الگوریتمی برای انتخاب ویژگی‌های مناســـب برای پیش‌بینی شـاخص، تحت عنوانISF_MID پیشـنهاد شـده و با تعدادی از روش‌های مشـابه، مقایسـه می‌گردد. داده‌های مورد اســـتفاده در این پژوهش به صـــورت روزانه در بازه زمانی ۱۳۹۲/۱۰/۲۸ تا ۱۳۹۷/۵/۳۰ جمع‌آوری شـده‌اند. مدل‌های مورد بررسـی در مرحله پیاده‌سـازی با روش اعتبارسـنجی متقابل K-fold مورد ارزیابی قرار گرفتند. همچنین از معیارهای MAE، MSE و RMSE برای ارزیابی عملکرد مدل‌های مذکور استفاده می‌شود. نتایج نشان می‌دهد که با روش پیشنهادی، می‌توان با ۷ ویژگی انتخابی به دقت بالایی در پیش‌بینی روزانه شاخص کل بورس اوراق بهادار تهران دست‌یافت.

کلیدواژه‌ها: پیش‌بینی شـــاخص بورس؛ الگوریتم انتخاب ویژگی؛ تابع پایه شـــعاعی؛ رگرسیون بردارپشتیبان؛ شبکه عصبی عمیق.

پرتال جامع علوم انسانی

تاریخ دریافت مقاله: ۱۳۹۹/۱۲/۰۳، تاریخ پذیرش مقاله: ۱۴۰۰/۰۶/۰۹.

* دانشجوی دکتری مدیریت مالی، دانشگاه شهید بهشتی، تهران، ایران.
** دانشیار، گروه مدیریت مالی و بیمه، دانشگاه شهید بهشتی، تهران، ایران (نویسنده مسئول) ،
E-Mail: m-fadaei@sbu.ac.ir
*** استاد، گروه مدیریت بازرگانی، دانشگاه شهید بهشتی، تهران،ایران.

۱. مقدمه

بازارهای مالی یکی از جذاب‌ترین ابداعات عصر حاضر می‌باشد. در این نوع بازارها، موفقیت یک سرمایه‌گذار به کیفیت اطلاعات و سرعت تصمیم‌گیری وی بستگی دارد. بنابراین پیش‌بینی آن‌ها برای سرمایه‌گذاران بسیار مهم است، زیرا پیش‌بینی دقیق به آن‌ها این امکان را می‌دهد آگاهانه درباره سرمایه‌گذاری‌ها تصمیم‌گیری کنند و ریسک‌ها را کاهش دهند. طبق نظریات گشت تصادفی[1] و بازار مالی کارا[2] هیچ سرمایه‌گذاری با استفاده از داده‌های تاریخی و فعلی نمی‌تواند سودی بیشتر از نرخ متوسط بازار بدست آورد، بنابراین امکان پیش‌بینی وجود ندارد[۳۱]. با این حال، پژوهشگران و سرمایه‌گذاران با فرضیه بازار کارآمد هم از لحاظ نظری و هم از لحاظ تجربی مخالف می‌باشند و معتقدند که متغیر های مالی مانند قیمت سهام، ارزش شاخص‌های بازار سهام و قیمت مشتقات مالی تا حدودی قابل پیش‌بینی هستند[۹]. بنابراین به ارائه و آزمایش مدل‌های مربوط به رفتار بازار سهام علاقه و توجه نشان داده و تئوری‌های متعددی در مورد بازارسهام مطرح کرده‌اند. در این میان محبوبیت شبکه‌های عصبی به‌دلیل رفتار بی‌نظم، بی‌ثبات، نویزی و ناپارامتریک بازارهای مالی رو به افزایش است، زیرا از توانایی ذاتی در تخمین روابط غیرخطی با دقت بالا، تحلیل‌های چند متغیره بدون هیچ فرض از پیش تعیین شده‌ای، و سهولت تعمیم برخوردارند که این امر باعث موفقیت نسبی آن‌ها در مدل‌سازی و پیش‌بینی بخش‌های متعدد بازارهای مالی شده است[۲۵].

از طرفی شبکه‌های عصبی در صورتی قادر به کشف درست اطلاعات پنهان هستند که ورودی‌های مناسبی انتخاب گردد. بنابراین، انتخاب ورودی‌های درست یکی از عوامل اصلی تأثیرگذار در کارایی شبکه‌های عصبی می‌باشد. براساس دیدگاه‌های ژونگ و انکه[3] (۲۰۱۷)، بازار سهام تحت تاثیر عوامل متعددی است که خود رابطه‌ای درون پیوسته با هم دارند. این عوامل شامل متغیرهای اقتصادی، سیاسی، روانشناختی و متغیرهای خاص شرکتی می‌شوند. بنابراین در بازار سهام صدها متغیر تأثیرگذار وجود دارد که برای تحلیل بازار مورد استفاده قرار می‌گیرند. تعداد بسیار زیادی از این متغیرها برای پیش‌بینی نامناسب واضافی هستند و یا این‌که کارایی چندانی ندارند. استفاده از متغیرهای نامرتبط منجر به سردرگمی روش هوش مصنوعی در بدست آوردن رابطه بین داده‌ها و درنتیجه عدم یادگیری صحیح شبکه می شود و علاوه بر مختل کردن یادگیری شبکه های عصبی، زمان یادگیری را نیز افزایش می‌دهد. بنابراین کاهش تعداد ویژگی‌هایی که باید در آموزش مدل به کار گرفته شوند، تأثیر به سزایی در افزایش دقت نتیجه و کاهش هزینه دارد. جهت کمینه کردن اثر وجود متغیرهای اضافی و نادرست در یادگیری شبکه‌های عصبی، از میان تمامی متغیرها لازم است آن‌هایی را انتخاب نمود که دارای ارتباط

---

[1] Random Walk Hypothesis
[2] Efficient Market Hypothesis
[3] Zhong and Enke

معناداری با خروجی باشـــند. بنابراین مطالعه در مورد متغیرهای تأثیرگذار بر بازدهی ســـهام و سنجش همبستگی و روابط میان متغیرها از اساسی‌ترین موضوعات حوزه مالی تجربی بالاخص درخصوص چگونگی قیمت‌گذاری سهام، توسعه استراتژی‌های پوشش ریسک، بهینه سازی پرتفوی و مدل‌سازی مالی است[۷]،که این امر می‌تواند بخش عمده‌ای از نیاز سرمایه‌گذاران و سهام‌داران را برآورده سازد. یکی از بهبودهای به‌کار رفته در مطالعات جدید، با توجه به پیشرفت کاربرد روش‌های هوش مصنوعی، استفاده از انتخاب ویژگی به عنوان یک پیش مرحله برای مدل پیش‌بینی می‌باشـــد، زیرا همان‌طور که اشـــاره گردید، دقت عملکرد مدل مورد اســـتفاده برای پیش‌بینی به میزان قابل توجهی تحت تأثیر تعداد ویژگی‌های ورودی آن است.

این پژوهش با هدف ارائه الگوریتمی برای انتخاب ویژگی‌های مناســـب به منظور پیش‌بینی روزانه شاخص کل بورس اوراق بهادار تهران با کمترین خطا، بر روی سه موضوع تمرکز می‌کند: ۱- ارائه یک روش مناسب برای اولویت‌بندی مجموعه‌ی ویژگی‌های ورودی مدل پیش‌بینی، ۲- مقایســـه دقت پیش‌بینی مدل‌های مختلف مورد بررســـی و انتخاب مدل مناســـب، ۳- انتخاب ویژگی‌های مناسـب برای مدل پیش‌بینی شـاخص کل بورس اوراق بهادار تهران. در این راسـتا؛ ابتدا ویژگی های مؤثر بر شـــاخص را با روش های تخمین کمینه‌ی افزونگی بیشـینه‌ی وابسـتگی[1](mRMR) اولویت‌بندی نموده و با مقایســه عملکرد دو روش، روش مناسـب انتخاب می‌گردد. سپس به‌منظور انتخاب مدل مناسب برای پیش‌بینی شاخص بورس از چهار مدل مختلف شـبکه عصـبی مصـنوعی، که از مهم‌ترین و بدیع‌ترین مدل‌های پیش‌بینی می‌باشـند، اسـتفاده می شود. ضمن مقایسه میزان دقت پیش‌بینی مدل‌ها نسبت به یکدیگر و انتخاب مدل مناسب، اقدام به شناسایی ویژگی‌های موثر در کاهش خطای مدل پیش‌بینی می‌شود. در نهایت الگوریتمی برای انتخاب ویژگی‌های مناسب به‌عنوان ورودی‌های مدل پیش‌بینی روزانه شاخص بورس اوراق بهادار تهران معرفی می‌گردد. بنابراین مهم‌ترین دســـتاورد این پژوهش، ارائه یک رویکرد نوین برای انتخاب ویژگی‌های مناسـب مدل پیش‌بینی با هدف فیلتر کردن متغیرهای ورودی بی‌ربط و اضافی برای کاهش پیچیدگی و درنتیجه بهبود دقت مدل پیش‌بینی شاخص بورس می‌باشد.

۲. مبانی نظری و پیشینه پژوهش

مبانی نظری

پیش‌بینی بازار سهام موضوعی کلاسیک می‌باشد که در محافل مالی و دانشگاهی مطرح است. طبقه‌بندی‌های مختلفی برای روش‌های پیش‌بینی بازارهای سهام وجود دارد، که عبارتند از:

الف) رویکرد تکنیکی

از اوایل قرن بیستم، رفتار قیمت سهام و ارزش آن به شکلی علمی‌تر مورد توجه قرار گرفت، برخی از تحلیل‌گران و شرکت‌های سرمایه‌گذاری از طریق تعقیب قیمت و روندهای خاص، الگوی

[1] Minimum Redundancy Maximum Relevance

تغییرات قیمت را بدست آورده و نتایج کارهای خود را مبنای تصمیمات سرمایه‌گذاری قرار می‌دادند. ترسیم رفتار قیمت، بررسی و تهیه نمودارها و مطالعه نوسانات و شناخت حساسیت‌های رفتار قیمت و پیش‌بینی آینده آن، هدف اصلی این گروه از صاحب‌نظران می‌باشد. این گروه را تحلیل‌گران تکنیکی یا چارتیست[1] می‌خوانند، زیرا از منحنی‌ها و نمودارها، استفاده زیادی به‌عمل می‌آورند. آن‌ها اعتقاد دارند که قیمت‌های گذشته، منعکس‌کننده آینده بوده و قیمت را تابع محض عرضه و تقاضا می‌دانند و معتقدند که عوامل موثر بر عرضه و تقاضا بی‌شمارند و هیچگاه نمی‌توان آن‌ها را به درستی و دقت شناسایی نمود، لذا بهترین شیوه‌کار را مطالعه حرکات گذشته و بدست آوردن الگوی تغییرات آینده می‌دانند. آنان به دنبال تغییرات بلندمدت نیستند و به دنبال حداکثر استفاده از فرصت‌های کوتاه‌مدت و بدست آوردن سود آنی هستند. این روش در حال حاضر علاقه‌مندان بسیار زیادی درمیان تحلیل‌گران بازار دارد. برخی از حوزه‌های تجزیه‌وتحلیل فنی عبارتند از: احساسات[2]، داده‌های خام[3]، روند[4]، مومنتوم[5]، حجم[6]، و نوسانات[7]. احساسات بیانگر رفتارهای شرکت‌کنندگان در بازار است. داده‌های خام شامل قیمت‌های سهام و الگوهای قیمت مانند نمودار خطی، نمودار شمعی و نمودار میله‌ای می‌باشند. روند و مومنتوم مثال‌هایی از شاخص‌های مبتنی بر قیمت هستند، روند برای ردیابی قیمت بازار سهام به‌کار می‌رود، در حالی‌که مومنتوم برای ارزیابی سرعت تغییر قیمت و قضاوت در مورد این‌که آیا روند معکوس قیمت سهام در شرف وقوع است، استفاده می‌شود. حجم شاخصی است که اشتیاق خریدار و فروشنده را برای سرمایه‌گذاری نمایان می‌سازد، همچنین مبنایی برای پیش‌بینی تغیرات قیمت سهام است، و نوسانات که اغلب برای بررسی دامنه نوسان قیمت سهام و ارزیابی ریسک و شناسایی سطح پایداری و مقاومت استفاده می‌شود[۲۰].

ب) رویکرد بنیادی

این رویکرد و مدل‌های مورد استفاده در آن از دهه ۱۹۳۰ مطرح بوده، اما عمدتا بعد از جنگ جهانی دوم در قالب نظری فراگیر، مورد توجه قرار گرفتند. در این‌گونه مدل‌ها اساسا به ارزش ذاتی[8] سهم توجه می‌شود که در واقع همان به‌کارگیری تمامی اطلاعات موجود و در دسترس است. در این روش وضعیت سهام یک شرکت را با توجه به تمام ابعاد اصلی خرد و کلان، مورد توجه و ارزیابی قرار می‌دهند. براساس نظر هیو وهمکارانش[9](۲۰۱۵)، اصولا تجزیه و تحلیل

[1] Chartist
[2] Sentiment
[3] Raw data
[4] Trend
[5] Momentum
[6] Volume
[7] Volatility
[8] Intrinsic Value
[9] Hu et al.

بنیادی در سه سطح اصلی انجام می شود: ۱- تجزیه و تحلیل اقتصادکلان که تاثیرات محیط اقتصادکلان و سیاست‌های پولی و مالی دولت را بر سود آینده یک شرکت مورد تحلیل قرار می‌دهند. ۲- تجزیه و تحلیل صنعت که ارزش شرکت را براساس وضعیت وچشم انداز صنعت، برآورد می‌کنند. ۳- تجزیه و تحلیل شرکت که عملکرد فعلی و وضعیت مالی شرکت را برای ارزیابی ارزش ذاتی آن، بررسی می‌کنند.

ج) رویکرد مبتنی بر نظریه‌های مدرن مالی

پس از جنگ جهانی دوم، مجموعه نظریه‌های جدیدی به جامعه مالی عرضه شد که عنوان نظریه‌های مدرن مالی را به‌خود گرفت. فرضیه بازار کارا، نظریه پرتفوی و مدل قیمت‌گذاری دارایی‌های سرمایه‌ای (CAPM) از حوزه‌های این رویکرد می‌باشند. این نظریه‌ها بر بازارهای مالی جهان تاثیری عمیق گذاشته‌اند و پایه فکری و مرکز ثقل تصمیمات رد یا قبول سرمایه‌گذاری برای بسیاری از پروژه‌ها، در نسلی از مدیران مالی و سرمایه‌گذاران بوده‌اند. اما تحولات بورس اوراق بهادار نیویورک در سال‌های ۱۹۸۷ به شدت اعتبار فرضیه بازار کارا و مدل‌هایی مانند تصادفی بودن قیمت را مورد سوال قرار داد[۴۹]. بنابراین تناقض‌های آشکار میان نظریه‌های موجود و دنیای واقعی موجب شده‌است تا دانشگاهیان و اهل نظر، نسبت به اعتبار نظریه‌های مدرن مالی به دیده تردید بنگرند.

د) رویکرد بی‌نظمی و پویایی غیرخطی

از اواسط دهه ۱۹۷۰ و به‌ویژه از سال ۱۹۹۰ کوشش‌های جدید و گسترده‌ای در زمینه پیش‌بینی قیمت سهام، با استفاده از روش‌های جدید ریاضی، سری‌های زمانی طولانی و تکنیک‌های هوش مصنوعی آغاز گردید که منجر به ظهور دیدگاه بی‌نظمی و پویایی غیرخطی شد. اندیشمندان در این چارچوب، در جستجوی روش‌های علمی جدیدتری هستند و معتقدند، نظریه بی‌نظمی و پویایی غیرخطی می‌تواند مقدمه ارائه نظریه‌های علمی تازه‌ای در مدیریت مالی باشد. گروهی از صاحب‌نظران و اندیشمندان، قدم فراتر گذاشته و به پدیده آشفتگی[1] توجه نمودند. این گروه مانند ریاضی‌دانان و فیزیک‌دانان، اعتقاد دارند چنانچه هر پدیده‌ای به درستی مشاهده و بررسی شود، وقایع به ظاهر تصادفی آن تبیین‌پذیر است و اگر در خور پیش‌بینی نباشد، حداقل قابل درک خواهد بود.

طرفداران الگوی آشفتگی اذعان می‌کنند که رفتار قیمت‌ها غیرخطی است و حرکات آن ناشی از عوامل بسیار زیادی است که در هر لحظه از زمان ممکن است بر آن تاثیر بگذارند. آن‌ها معتقدند که به تعداد سرمایه‌گذاران، عامل تاثیرگذار بر قیمت سهام وجود دارد. در این راستا به‌دلیل

[1] Chaos

وجود روابط غیرخطی پیچیده با متغیرهای متعدد، توانایی و قابلیت مدل‌های رگرسیون خطی به‌دلیل زمان‌بر بودن آن‌ها در پردازش اطلاعات کاهش می‌یابد و لذا باید از یک سری روابطی مبتنی بر شبکه‌های عصبی استفاده شود. زیرا در شبکه‌های عصبی، شبیه سازی رفتار بازار با سرعت زیادتر و دقت بالاتری صورت گرفته و درنتیجه توانایی در تصمیم‌گیری را بهبود می‌بخشد[۵۱]. بنابراین شبکه‌های عصبی مصنوعی با تحلیل‌های چند متغیره با ویژگی غیرخطی، داده‌های استنتاجی و سهولت تعمیم، الگوریتم‌های پیشتاز و اصلی در استفاده از روش‌های یادگیری ماشین در مطالعات مالی بوده و تبدیل به یک ابزار عامه پسند و غالب در امور مالی و اقتصاد شده‌اند.

باتوجه به این‌که عملکرد شبکه‌های عصبی، به‌طور مستقیم با کیفیت ویژگی‌ها مرتبط است و قرار دادن تعداد بیش از اندازه متغیر مستقل در رگرسیون، یکی از بزرگ‌ترین موانع در کاربرد عملی آن‌ها، می‌باشد. بنابراین برای پیش‌بینی شاخص بورس، ویژگی‌هایی حاوی اطلاعات مفید مورد نیاز است. از آنجائی‌که حرکات قیمت سهام، نتایج چندین عامل از جمله اقتصاد کلان، وضعیت صنعت، عملکرد مالی شرکت‌ها، احساسات سرمایه گذاران و غیره است، انتخاب ویژگی‌های مفید نقش مهمی در پیش‌بینی شاخص بورس دارند. انکه و تاورن[1] (۲۰۰۵)؛ نشان می‌دهند که طراحی یک شبکه عصبی مصنوعی با حداقل پیچیدگی و متغیرهای ورودی تاثیرگذار و کاملا مرتبط می‌تواند کارایی و دقت پیش‌بینی سری‌های زمانی مالی را بهبود بخشد. لام[2] (۲۰۰۴)، بیان می‌کند که برای شبکه‌های عصبی مصنوعی دستیابی به نتایج دقیق، گزینش تعمدی متغیرهای ورودی و یک ترکیب بهینه از پارامترهای شبکه، حیاتی است. آتسالاکیز و والاوانیز[3] (۲۰۰۹)، بیان می‌کنند که در میان مطالعات معطوف به پیش‌بینی بازار سهام برخی از محققان به پیش‌پردازش ساده‌ای از داده‌ها، با استفاده از تبدیل آن‌ها به داده‌های لگاریتمی یا نرمال‌سازی داده‌های خام، اقدام می‌کنند. درحالی‌که دیگران داده‌ها را پیش‌پردازش نکرده یا جزئیات بیشتری درباره پالایش داده‌ها ارائه نمی‌کنند، در صورتی‌که کلید موفقیت پیش‌بینی بازار سهام، دستیابی به بهترین نتایج با کم‌ترین داده‌های ورودی موردنیاز و کم‌ترین پیچیدگی در مدل پیش‌بینی می‌باشد. بنابراین اگر از یک شبکه عصبی مصنوعی انتظار می‌رود تا یک پیش‌بینی دقیق و کارآمد را ارائه دهد، ضروری است تا ورودی‌های بسیار موثر و معرف[4] انتخاب شوند. این نوع گزینش وظیفه اصلی تکنولوژی کاهش ابعاد[5] می‌باشد. محققین در علوم کامپیوتر، آمار و ریاضیات کاربردی در این زمینه سال‌های زیادی کار کردند و انواع مختلفی از تکنیک‌های کاهش ابعاد خطی و غیرخطی را توسعه دادند. وندرمتن و همکاران[6] (۲۰۰۹)، یک مرور و مقایسه

[1] Enke &Thaworn
[2] Lam
[3] Atsalakis & Valavanis
[4] Representive
[5] Dimensionality Reduction
[6] Van Der Maaten et al.

سیستماتیک از این تکنیک‌ها را ارائه می‌دهند. همچنین سورزانو و همکاران[1] (۲۰۱۴)، بسیاری از تکنیک‌های کاهش ابعاد را با بینش ریاضی پشت آن‌ها، طبقه‌بندی می‌کنند. یکی از شیوه‌های کاهش ابعاد انتخاب مرتبط‌ترین متغیرها از مجموعه داده‌های اصلی می‌باشد که انتخاب ویژگی نامیده می شود. برای انتخاب ویژگی، راه‌حل‌ها و الگوریتم‌های فراوانی ارائه شده ا ست. م شکل بعضی از الگوریتم‌های ارائه شده، بار محاسباتی زیاد آن‌ها می‌با شد. بنابراین؛ همچنان پیدا کردن یک الگوریتم سریع برای انتخاب ویژگی بسیار مهم می‌باشد.

پیشینه پژوهش

روند مطالعات در پژوهش‌های داخلی و خارجی، برتری انواع شبکه‌های عصبی مصنوعی را به سایر تکنیک‌های محاسباتی هوشمند و نیز روش‌های آماری نشان می‌دهد. هنریکه، سوبریرو و کیمورا[2] (۲۰۱۹)، ۵۴۷ مقاله را بررسی کرده و بیان نمودند که حدود ۷۴ درصد از مطالعات بررسی شده از نوعی شبکه عصبی مصنوعی برای پیش‌بینی بازار سهام استفاده کردند. در ادامه به ذکر چند نمونه از مطالعاتی پرداخته می‌شود که بیشترین ارتباط را با موضوع پژوهش دارند.

ذوالفقاری، سحابی و بختیاران(۱۳۹۹)، به‌منظور پیش‌بینی روزانه بازدهی شاخص کل بورس اوراق بهادار تهران، مدلی ترکیبی از شبکه یادگیری عمیق و مدل‌های منتخب خانواده GARCH ارائه نمودند، در ساختار این مدل از میان مدل‌های شبکه یادگیری عمیق، شبکه عصبی بازگشتی مبتنی بر حافظه کوتاه‌مدت و بلندمدت (RNN-LSTM) و از مدل‌های دارای حافظه‌کوتاه‌مدت، GARCH و EGARCH استفاده شده است. همچنین از قیمت نفت و نرخ دلار به‌عنوان متغیر م ستقل بهره برده شده ا ست. نتایج تحقیق ن شان می‌دهد که مدل‌های ترکیبی دقت پیش‌بینی بالاتری نسبت به مدل‌های تکی دارند، و مدل ARMAX-EGARCH-RNN-GED نسبت به ۲۳ مدل دیگر ترکیبی، خطای پیش‌بینی کمتری دارد (RMSE=۰,۳۳).

صالحی و گرشاسبی(۱۳۹۸)، یک روش نوین ترکیبی برای پیش‌بینی شاخص کل بورس اوراق بهادار تهران ارائه کردند که همزمان از الگوریتم رقابت استعماری به‌عنوان روش انتخاب ویژگی و شبکه عصبی فازی انطباق‌پذیر به‌عنوان تابع پیش‌بینی‌کننده استفاده شده است. برای ورودی مدل از ۶۸ ویژگی مؤثر بر بورس اوراق بهادار بهره برده شده است. با ۸۰ دفعه اجرای مدل تعداد ۱۰ ویژگی انتخاب شدند که بیشترین تأثیر را در پیش‌بینی شاخص‌کل بورس می‌گذارند. طبق نتایج مشخص گردید، شاخص‌های اقتصادی بیشترین تأثیر و شاخص بورس بازارهای بین‌المللی کمترین تأثیر را دارند.

تهرانی، حیرانی و من صوری (۱۳۹۸)، در پژوه شی با عنوان "مقایسه عملکرد مدل پنج عاملی فاما و فرنچ و انواع رویکردهای شبکه عصبی و عصبی فازی در پیش‌بینی قیمت سهام"، به مقایسه دقت مدل ARMA، انواع شبکه‌های عصبی مصنوعی و شبکه‌های عصبی فازی در

[1] Sorzano et al.
[5] Henrique, Sobreiro & Kimura

پیش‌بینی قیمت سهام در بورس اوراق بهادار تهران پرداختند. نتایج بدست آمده با ۶ معیار ارزیابی دقت پیش‌بینی، نشان می‌دهد که شبکه عصبی تابع پایه شعاعی، شبکه عصبی رگرسیون بردار پشتیبان، شبکه عصبی فازی و شبکه عصبی چندجمله‌ای، به‌ترتیب نسبت به دیگر مدل‌های شبکه‌های عصبی و ARMA از دقت بالاتری برخوردار می‌باشند.

باجلان، فلاح‌پور و دانا(۱۳۹۶)؛ مدلی برپایه‌ی ماشین بردار پشتیبان وزن‌دهی شده همراه با روش انتخاب ویژگی هیبرید که مرکب از یک بخش فیلترکننده و یک بخش پوشش‌دهنده به‌منظور انتخاب زیرمجموعه‌ای بهینه از ویژگی‌ها می‌باشد، معرفی نمودند. براساس نتایج، مشخص گردید مدل VW-SVM همراه با انتخاب ویژگی F-SSFS عملکرد بهتری در پیش‌بینی قیمت سهم، نسبت به روش‌های موجود دارد.

راعی، نیک عهد قصیرائی و حبیبی (۱۳۹۵)، در پژوهشی تحت عنوان "پیش‌بینی شاخص بورس اوراق بهادار تهران با ترکیب روش‌های آنالیز مولفه‌های اصلی، رگرسیون بردار پشتیبان و حرکت تجمعی ذرات"، به‌منظور افزایش دقت پیش‌بینی شاخص بورس اوراق بهادار تهران ابتدا از روش PCA برای پالایش اولیه داده‌ها استفاده کرده و سپس با استفاده از رگرسیون بردار پشتیبان بهینه شده به‌و سیله الگوریتم حرکت تجمعی ذرات، به پیش‌بینی شاخص اقدام نمودند. نتایج بدست آمده نشان می‌دهد که پیش‌پردازش داده‌ها، خطای پیش‌بینی مدل را به‌طور قابل توجه‌ای کاهش می‌دهد. سینگ و سریواستاوا[1] (۲۰۱۷)، و ژونگ و انکه[2](۲۰۱۷) نیز از تکنیک PCA برای کاهش ابعاد ویژگی‌ها استفاده نمودند و به نتایج مشابهی دست یافتند.

حقیقت منفرد، احمدعلی‌نژاد و متقالچی (۱۳۹۱)، به بررسی و مقایسه دقت سه مدل از شبکه های عصبی (پرسپترونی چندلایه، پایه‌ای شعاعی و رگرسیونی) و یک مدل از مدل‌های سری زمانی (باکس–جنکینز) در پیش‌بینی شاخص کل قیمت بورس تهران پرداختند. نتایج حاصله حاکی از آن است که ۳ مدل شبکه عصبی به کار رفته از لحاظ ۴ معیار خطا نسبت به مدل سری زمانی برتری دارند. از طرفی از میان ۳ مدل شبکه عصبی به کار رفته به‌ترتیب، مدل شبکه عصبی پایه شعاعی و پس از آن مدل شبکه عصبی پروسپترون چند لایه بهترین عملکرد و شبکه عصبی رگرسیونی بدترین عملکرد را ارائه کرده‌اند. این نتایج با مطالعات پاکدین امیری و همکاران(۱۳۸۸)، تهرانی و مرادپور(۱۳۹۱)، و گوئرسن و همکاران[3] (۲۰۱۱) مطابقت دارد.

فلاح‌پور، گل‌ارضی و فتوره‌چیان (۱۳۹۲)، به‌منظور پیش‌بینی تغییرات قیمت سهام در بورس تهران، متغیرهای ورودی ماشین بردار پشتیبان را توسط الگوریتم ژنتیک، بهینه سازی نمودند. نتایج نشان داد که ماشین بردار پشتیبان برپایه الگوریتم ژنتیک، دقت بسیار بیشتری نسبت به ماشین بردار پشتیبان ساده دارد.

---

[1] Singh & Srivastava
[2] Zhong & Enke
[3] Guresen et al.

هوانگ[1] (۲۰۱۲)، به‌منظور بهینه‌سازی پارامترهای رگرسیون بردار پشتیبان و انتخاب ورودی‌های مدل از الگوریتم ژنتیک استفاده کرد. همچنین چی‌یو و سونگ[2](۲۰۱۶)، به‌منظور پیش‌بینی یک روز بعد جهت قیمت شاخص بازار سهام ژاپن از یک شبکه عصبی مصنوعی بهینه شده با الگوریتم‌های ژنتیک استفاده نموده و برای ارزیای تاثیر نوع متغیر ورودی بر عملکرد مدل، دو نوع اساسی از متغیرهای ورودی را مورد بررسی قرار دادند. نتایج تجربی نشان داد که با انتخاب مناسب متغیرهای ورودی، می‌توان عملکرد شبکه‌های عصبی مصنوعی را بهبود بخشید.
الحق و همکاران[3] (۲۰۲۱)، به منظور پیش‌بینی روند سهام، ویژگی‌های انتخاب شده توسط ماشین بردار پشتیبان، رگرسیون لجستیک منظم و جنگل تصادفی را برای ایجاد زیر مجموعه‌ای بهینه از ویژگی‌ها ترکیب کرده و سپس از یک مدل سازنده عمیق برای پیش‌بینی حرکات آینده قیمت سهام ا ستفاده نمودند. برا ساس یافته‌ها، به این نتیجه ر سیدند که اثربخ شی یک رویکرد ترکیبی انتخاب ویژگی در بهبود عملکرد یک مدل مولد عمیق بیشتر می‌باشد.
لانگ، لو و کوئی[4](۲۰۱۹)، مدلی جدید تحت عنوان شبکه عصبی چند فیلتری را برای استخراج ویژگی‌ها، معرفی کردند. برای ایجاد ساختار چند فیلتری هر دو نورون عصبی کانولوشنال و بازگشتی یکپارچه شده و سپس برای پیش‌بینی شاخص بازار سهام چین مورد استفاده قرار گرفتند. برا ساس نتایج م شخص گردید که مدل پی شنهادی از مدل‌های سنتی یادگیری ما شین، آماری و تک ساختاری (کانولوشنال، بازگشتی و حافظه کوتاه‌مدت بلندمدت) از نظر دقت، سودآوری و ثبات بهتر عمل می‌کند.
گوندوز، کتلتپه[5] (۲۰۱۵)، از یک روش انتخاب ویژگی به نام اطلاعات متقابل متعادل[6] (BMI) به‌منظور شناسایی ویژگی‌های مرتبط‌تر برای تعیین جهت بازار استفاده نمودند و سپس یک الگوریتم بیز ساده را برای مدل سازی بردارهای ویژگی و قیمت‌های سهام، و پیش‌بینی حرکات آینده بازار به‌کار بردند.
گریگوریان[7] (۲۰۱۶) برای پیش‌بینی شاخص‌های بورس اوراق بهادار بخار ست و بازار سهام بالتیک، از یک مدل پیش‌بینی مبتنی بر ماشین بردار پشتیبان و تکنیک تحلیل مؤلفه‌های م ستقل[8] (ICA)، با نام SVM-ICA ا ستفاده کرد. نتایج نشان می‌دهد که مدل پی شنهاد شده در پیش‌بینی سری‌های زمانی نامانا بهتر از مدل ماشین بردار پشتیبان عمل می‌کند.
همان‌طور که در پژوهش‌های ذکرشده اشاره گردید دو تکنیک برای کاهش ابعاد ویژگی‌ها برای یک مدل پیش‌بینی وجود دارد. یکی از تکنیک ها روش های استخراج ویژگی است.

[1] Huang
[2] Qiu & Song
[3] Ul Haq et al.
[4] Long, Lu & Cui
[5] Gunduz & Cataltepe
[6] Balanced Mutual Information
[7] Grigoryan
[8] Independent Component Analysis

مکانیسم‌های استخراج ویژگی سعی می‌کنند مجموعه ویژگی‌های اصلی را به فضای ویژگی‌هایی با ابعاد کمتر انتقال دهند، بدون این‌که ماهیت مسئله تغییر کند، اما در این انتقال معنای اصلی ویژگی‌ها تغییر می‌کند، بنابراین امکان اولویت‌بندی و شناسایی ویژگی‌های انتخابی وجود ندارد. اما تکنیک دیگر کاهش ابعاد یعنی انتخاب ویژگی که به آن انتخاب زیر مجموعه ویژگی نیز گفته می شود، ویژگی‌هایی را که برای مدل سازی داده‌های آموخته شده نامرتبط هستند، شناسایی و برخلاف روش‌های مبتنی بر استخراج ویژگی، معنای اصلی ویژگی‌ها را بعد از کاهش حفظ می‌کند[۸]. اغلب پژوهش‌های صورت گرفته با این تکنیک، از ترکیب الگوریتم‌های فرااکتشافی[1] مانند الگوریتم ژنتیک و شبکه‌های عصبی استفاده می‌کنند که چالش‌ها و معایبی را به‌همراه دارد که از جمله می‌توان به موارد ذیل اشاره نمود:

- این روش‌ها برای مجموعه‌داده‌هایی با تعداد نمونه‌ها و ویژگی‌های زیاد، بسیار زمان‌بر می‌باشد.
- با استفاده از این روش‌ها، تنها می‌توان به یک زیرمجموعه از مجموعه کل ویژگی‌ها دست یافت که مقدار خطای کمتری نسبت به دیگر زیرمجموعه‌های بررسی شده دارد. این زیرمجموعه، ممکن است یک بهینه محلی برای مسائل بهینه‌سازی باشد.
- در هنگام انتخاب مجموعه ویژگی‌های مناسب، وابستگی بین ویژگی‌های ورودی در نظر گرفته نشده و فقط از همبستگی ویژگی‌های ورودی نسبت به خروجی مدل استفاده می‌شود. در نتیجه، احتمال انتخاب زیرمجموعه‌ای با افزونگی، افزایش می‌یابد.
- همچنین، در این روش‌ها اولویت‌بندی ویژگی‌ها انجام نمی‌شود.

بنابراین با رویکرد پیشنهادی در مطالعه حاضر، سعی شده است راه‌حلی برای رویارویی با این چالش‌ها ارائه گردد. به‌طوری‌که، در این رویکرد نوین ضمن توجه به همبستگی هر ویژگی با خروجی مدل، همبستگی بین ویژگی‌های ورودی با یکدیگر نیز بررسی می‌شود. همچنین به اولویت‌بندی ویژگی‌ها پرداخته شده و تلاش شده است به‌منظور کاهش هزینه و بار محاسباتی، حداقل ویژگی‌های مناسب انتخاب شوند. این رویکرد نسبت به سایر روش‌های ارائه شده در این حوزه، باعث بهبود قابل توجهی در دقت مدل پیش‌بینی می‌شود. در مجموع می‌توان گفت متدولوژی ارائه شده در این پژوهش جهت شناسایی، اولویت‌بندی و انتخاب ویژگی‌های مناسب برای بهبود عملکرد مدل پیش‌بینی شاخص بورس بازگوکننده نوآوری این مقاله می‌باشد.

سوال‌های پژوهش

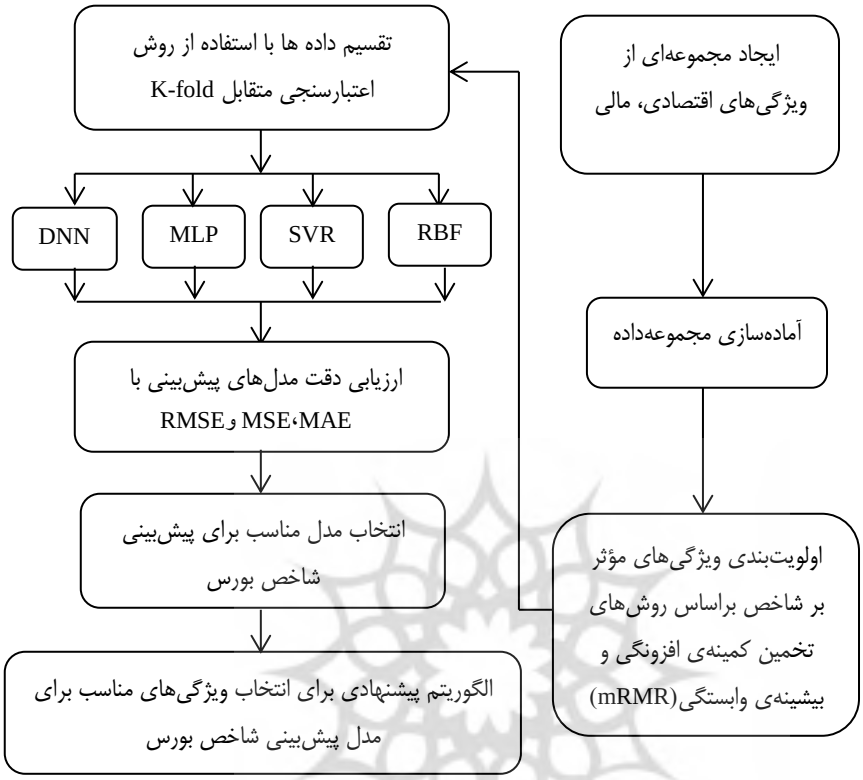
این پژوهش به‌دنبال پاسخگویی به سوالات زیر است:

۱. برای تخمین mRMR کدام یک از الگوریتم‌های انتخاب ویژگی با نام‌های MID و FCD، مناسب می‌باشند؟

[1] Metaheuristic

۲. اولویت ویژگی‌های انتخاب شده توسط MID و FCD، به چه صورت است؟
۳. کدام یک از مدل‌های پیش‌بینی شاخص بورس با نام‌های MLP،RBF، SVR و DNN دقت بیشتری در پیش‌بینی شاخص بورس اوراق بهادار تهران دارند؟
۴. ویژگی‌های مناسب برای مدل پیش‌بینی روزانه شاخص بورس اوراق بهادار تهران کدامند؟

برای پاسخگویی به سوالات فوق مراحل نمایش داده در شکل(۱)، انجام می‌شود.



شکل۱. مراحل انجام پژوهش

۳. روش‌شناسی پژوهش

در این بخش، مجموعه داده، تکنیک اولویت‌بندی ویژگی‌های مؤثر بر شاخص، مدل‌های مورد استفاده و معیارهای ارزیابی عملکرد مدل‌های پیش‌بینی، معرفی می‌شوند.

ایجاد مجموعه داده

داده‌های اصلی در این پژوهش، شاخص‌های مالی و اقتصادی استخراج شده از مبانی تئوریک که در مطالعات ژونگ و انکه (۲۰۱۷)، اسدی و همکاران (۲۰۱۲)، پشوتنی‌زاده و همکاران(۱۳۹۹)، فقیهی‌نژاد و مینایی (۱۳۹۷)، و بدیعی, رضـــازاده و محمودی (۱۳۹۶)، مورد تائید واقع شـــده‌اند، می‌باشـــند. این داده‌ها در جدول ۱ شـــرح داده شـــده‌اند. با توجه به‌اینکه افزایش تعداد داده‌ها در

شبکه‌های عصبی مصنوعی موجب کسب نتایج دقیق‌تر می شود، لذا داده‌های مورد استفاده در پژوهش بصـــورت روزانه از تاریخ ۱۳۹۲/۱۰/۲۸ تا ۱۳۹۷/۵/۳۰ به‌مدت ۱۱۰۸روز کاری بورس تهران جمع‌آوری شده‌اند.

جدول ۱. فهرست ویژگی‌های مورد بررسی برای پیش‌بینی شاخص کل بورس اوراق بهادار تهران

شاخص	توصیف مختصر ویژگی	نام ویژگی
شاخص‌های مالی	بازده شاخص کل در سه روز گذشته	Tehran Exchange Dividend and Price Index (TEDPIX)
	تغییر نسبی حجم معاملات درسه روز گذشته	Trading Volume (TV)
	بازده شاخص ۵۰ شرکت فعال‌بورس درسه روز گذشته	Index of 50 More Active Companies (IC50)
	بازده شاخص قیمت ۵۰ شرکت فعال‌بورس درسه روز گذشته	Price Index of 50 More Active Companies (PIC50)
	بازده شاخص ۳۰ شرکت بزرگ درسه روز گذشته	30 largest Companies Index (LCI30)
	بازده شاخص صنعت درسه روز گذشته	Industrial Index (InI)
	بازده شاخص مالی درسه روز گذشته	Financial Index (FI)
	میانگین متحرک نمایی قیمت پایانی ۱۰، ۲۰، ۵۰ و ۲۰۰روز گذشته	Exponential Moving Average (EMA)
	میانگین متحرک قیمت پایانی ۵، ۱۰ ، ۳۰ روز گذشته	Moving Average (MA)
	تفاوت نسبی در درصد بازده شاخص در ۵، ۱۰،۱۵ و ۲۰ روز گذشته	Relative Difference in Percentage of the TEDPIX Return (RDP)
	شاخص سطح مقاومت قیمت برای ۳۰ و ۱۲۰ روز گذشته	Moving Price level Percentage (MPP)
	نسبت قیمت به درآمد با ۴وقفه و تغییرات نسبی ۳ روز گذشته	Price /Earning per Share (P/E)
شاخص‌های اقتصادی	قیمت نفت اوپک با ۴ وقفه و تغییرات نسبی ۳ روز گذشته آن	OPEC Oil Prices (OP)
	قیمت دلار با ۴ وقفه و تغییرات نسبی ۳ روز گذشته آن	USD/IRR (DP)
	قیمت یورو با ۴ وقفه و تغییرات نسبی ۳ روز گذشته آن	Euro/IRR (EP)
	قیمت سکه طلا با ۴ وقفه و تغییرات نسبی ۳ روز گذشته آن	Gold Coin Price (GP)

آماده‌سازی مجموعه‌داده

برای یکسان کردن مقیاس پارامترها، مرحله‌ی نرمال‌سازی انجام شده است. برای نرمال‌سازی پارامترها که شامل تعیین مقدار هر پارامتر در بازه‌ی بین صفر و یک است، از رابطه ۱ استفاده می‌شود:

$$V_{i_{norm}} = \frac{V_i - \min(V)}{\max(V) - \min(V)}$$ رابطه(۱)

که V_i و $V_{i_{norm}}$ به‌ترتیب مقدار اصلی و مقدار نرمال شده نمونه i ام متعلق به ویژگی V است. همچنین، $\min(V)$ و $\max(V)$ به‌ترتیب حداقل مقدار و حداکثر مقدار مربوط به نمونه‌های ویژگی V است.

۳-۴. اولویت‌بندی ویژگی‌های مؤثر بر شاخص

یکی از پیش‌پردازش‌های با اهمیت برای ایجاد یک مدل پیش‌بینی، انتخاب ویژگی‌های مناسب برای آن است. برای این منظور، روش‌های مختلفی اعم از الگوریتم‌های مبتنی بر هوش مصنوعی و الگوریتم‌های برگرفته از الگوهای آماری ارائه شده‌اند. الگوریتم‌های آماری در زمان کمتر نسبت به الگوریتم‌های هوشمند اجرا شده و از طرفی این امکان را دارند که برای هر یک از ویژگی‌ها اولویت مشخصی را تعیین کنند. یکی از مهم‌ترین رویکردها در اولویت‌بندی یک ویژگی، توجه به مقدار همبستگی آن با ویژگی هدف و همچنین با مجموعه ویژگی‌های انتخاب‌شده قبل از آن است. از روابط مورد استفاده برای تعیین مقدار همبستگی، می‌توان اطلاعات متقابل[1](MI) را نام برد که با استفاده از رابطه ۲ محاسبه می‌شود.

$$I(x;y) = \iint p(x,y)\log\frac{p(x)}{p(x)p(y)}dxdy$$ رابطه(۲)

در این رابطه $p(x)$ و $p(y)$ به‌ترتیب توابع چگالی احتمال متغیرهای x و y است. همچنین، $p(x,y)$، تابع چگالی احتمال وقوع هم‌زمان متغیرهای x و y می‌باشد.

در انتخاب مجموعه ویژگی‌های مناسب، نیاز است که به دنبال یافتن مجموعه ویژگی‌هایی (S) باشیم که بیش‌ترین مقدار وابستگی را با ویژگی هدف (h) داشته باشند. از رابطه ۳ برای این منظور استفاده می‌شود:

$$\max V_I(S,h),\quad V_I = \frac{1}{|S|}\sum_{i\in S} I(i,h)$$ رابطه(۳)

V_I نشان‌دهنده مقدار وابستگی مجموعه ویژگی‌های انتخابی با ویژگی هدف است. در این رابطه تنها به میانگین مقدار همبستگی مجموعه ویژگی‌های انتخابی با ویژگی هدف توجه شده است. در صورتی که همبستگی بین اعضای مجموعه ویژگی‌های انتخابی نیز اهمیت بالایی دارد. بر این اساس، از شرط حداقل افزونگی نیز در تعیین اولویت ویژگی‌ها استفاده می‌شود. در این شرط به دنبال زیرمجموعه‌ای از ویژگی‌ها هستیم که حداقل همبستگی را با یک‌دیگر داشته باشند. برای محاسبه این مقدار از رابطه ۴ استفاده می‌شود:

$$\min W_I(S),\quad W_I = \frac{1}{|S|^2}\sum_{i,j\in S} I(i,j)$$ رابطه(۴)

W_I میانگین MI بین ویژگی‌ها در مجموعه ویژگی‌های انتخابی است. با ترکیب دو رابطه ۳ و ۴، مفهومی تحت عنوان کمینه‌ی افزونگی بیشینه‌ی وابستگی[2] (mRMR) معرفی گردیده است[۳۴]. اساس mRMR با رابطه‌ای ۵ تعریف می‌شود:

$$\max \varphi(V,W),\quad \varphi = V - W$$ رابطه(۵)

که $\varphi(V,W)$ عملگری برای ترکیب روابط بیشینه‌ی وابستگی و کمینه‌ی افزونگی است. انتخاب

[1] Mutual Information
[2] Minimum Redundancy Maximum Relevance

لیســـت بهترین ویژگی‌ها برای مجموعه داده‌ای با تعداد زیادی ویژگی، به دلیل نمایی بودن تعداد محا سبه‌های لازم برای یافتن حداکثر مقدار $\varphi(V,W)$، منجر به غیرعملی شدن ا ستفاده از این تابع در یک زمان محدود می‌شـــود، به همین دلیل روش‌هایی برای تخمین mRMR معرفی گردیده‌اند. از متداول‌ترین آن‌ها، می‌توان به روش تفاضـــل اطلاعات متقابل[1] (MID) و اختلاف همبستگی آزمون F[2](FCD) اشاره کرد. MID با رابطه ۶ محاسبه می‌شود[۱۰]. باید توجه داشت که برای ا ستفاده از MID نیاز به گسسته سازی داده‌ها ا ست. برای این منظور با توجه به مقادیر بین کمینه و بیشینه‌ی شاخص بورس، تعداد بازه‌ها تعیین می‌شود.

رابطه(۶) $$max_{i\in\varphi(S)}[I(i,h)-\frac{1}{|S|}\sum_{j\in S}I(i,j)]$$

برای متغیرهای پیوسته از FCD استفاده شده که با رابطه‌ی ۷ محاسبه می‌شود[۲۷].

رابطه(۷) $$max_{i\in\varphi(S)}[F(i,h)-\frac{1}{|S|}\sum_{j\in S}c(i,j)]$$

که $c(i,j)$ مقدار همبستگی دو ویژگی i و j را نشان می‌دهد. تابع F با رابطه ۸ تعریف شده‌است.

رابطه(۸) $$F(g_i,h)=[\sum_k n_k(\bar{g}_k-\bar{g})/(K-1)]/\sigma^2$$

مقدار تابع F برای متغیر g_i، با توجه به K دسته‌ای که بر اساس h تعیین شده، محاسبه می‌شود. $\bar{g}$ میانگین مقدار g_i در همه نمونه‌ها و $\bar{g}_k$ میانگین مقدار در دسته‌ی kام ا ست. همچنین σ^2 که واریانس تلفیقی است با رابطه‌ی ۹ محاسبه می‌شود.

رابطه(۹) $$\sigma^2=[\sum_k(n_k-1)\sigma_k^2]/(n-K)$$

که σ_k^2 و n_k واریانس و اندازه‌ی دسته‌ی kام است.

مدل‌های مورد استفاده برای پیش‌بینی شاخص بورس

به‌منظور انتخاب مدل مناســب جهت پیش‌بینی بازدهی روزانه شــاخص بورس از چهار مدل به نام‌های تابع پایه شعاعی[3] (RBF)، رگر سیون بردار پ شتیبان[4] (SVR)، شبکه ع صبی پر سپترون چند لایه[5](MLP) و شبکه ع صبی عمیق[6] (DNN)، ا ستفاده می شود. RBF، SVR و MLP از مدل‌های بسیار محبوب و متداول برای مسائل تخمین تابع می‌با شند که در پژوهش‌های صورت گرفته در ارتباط با پیش‌بینی شاخص بورس در عر صه داخلی و خارجی، بارها مورد ا ستفاده قرار گرفته‌اند، و عملکرد آن‌ها در افزایش دقت پیش‌بینی در مقای سه با سایر مدل‌های هو شمند مورد

[1] Mutual Information Difference
[2] F-Test Correlation Difference
[3] Radial Basis Function
[4] Support Vector Regression
[5] Multi-Layer Perceptron
[6] Deep Neural Network

سنجش و تایید قرار گرفته است. از جمله این پژوهش‌ها می‌توان به مطالعات تهرانی، حیرانی و منصوری (۱۳۹۸)، حقیقت منفرد، احمدعلی‌نژاد و متقالچی (۱۳۹۱)، تهرانی و مرادپور(۱۳۹۱)، اوه و وانگ[1] (۲۰۰۹)، و وانگا و ژائو[2](۲۰۲۰) اشاره نمود. DNN از بدیع‌ترین شبکه‌های عصبی مصنوعی می‌باشد که هدف اصلی آن مدل‌سازی داده‌های پیچیده دنیای واقعی به‌وسیله استخراج ویژگی‌های مناسب از اطلاعات مرتبط است. DNN در انجام کارهایی نظیر طبقه‌بندی(لی و همکاران،۲۰۰۹)، تشخیص گفتار (ژانگ و وو[3]،۲۰۱۳) و کاهش ابعاد (هینتون و سالاخودینوف[4]،۲۰۰۶) به‌طور موفقیت‌آمیز استفاده شده‌اند. همچنین برخی از رویکردهای یادگیری عمیق درباره مدل‌سازی سری‌های زمانی مالی در پژوهش چونگ، هان و پارک (۲۰۱۷)، هیرانشا و همکاران[5] (۲۰۱۸)، لانگ، لو و کوئی (۲۰۱۹) و الحق و همکاران (۲۰۲۱)، ارائه شده است با این حال، کاربرد روش‌های یادگیری عمیق در زمینه پیش‌بینی‌های مالی یک حوزه نسبتا کشف نشده باقی مانده است[۳۹]. یکی از چالش اصلی در این پژوهش بررسی عملکرد شبکه عصبی عمیق در مقایسه با سایر شبکه‌های عصبی سنتی در پیش‌بینی شاخص بورس ایران می‌باشد. در ادامه بصورت اجمالی مدل‌های مورد استفاده معرفی می‌گردند.

تابع پایه شعاعی (RBF)

شبکه عصبیRBF، به‌طور گسترده‌ای برای تخمین توابع چندبعدی غیرپارامتری با بهره‌گیری از مجموعه محدودی از داده‌های یادگیری استفاده شده است. این شبکه تنها با یک لایه مخفی قادر به تخمین توابع می‌باشد، و اگر از نورون‌های کافی در لایه پنهان برخوردار باشد، قادر است هر تابع پیوسته را با دقت بالا تقریب بزند. در این شبکه بین نورون‌های لایه ورودی و لایه پنهان وزنی وجود ندارد، و در نورون‌های لایه پنهان تنها از توابع پایه شعاعی به عنوان تابع فعال سازی استفاده می‌شود. ورودی شبکه، برداری از اعداد حقیقی به صورت زیر است:

$$X = [x_1\ x_2\ \dots\ x_n]$$

هر نورون لایه پنهان از تابع گاوسی برای پردازش ورودی‌ها استفاده می‌کند. این تابع برای نورون i با رابطه‌ی ۱۰ تعریف می شود:

$$h_i = \phi_i(X) = \exp\left(-\frac{\|X-\mu_i\|}{2\sigma_i^2}\right) \quad \text{رابطه(۱۰)}$$

μ_i و σ_i به‌ترتیب مرکز و شعاع نورون i است.

[1] Ou & Wang
[2] Wanga & Zhao
[3] Zhang & Wu
[4] Hinton & Salakhutdinov
[5] Hiransha et al.

در این شبکه از یک نورون خروجی استفاده شده است. نورون موجود در لایه‌ی خروجی، یک ترکیب خطی از توابع لایه‌ی پنهان ایجاد می‌کند. بین نورون‌های لایه‌ی پنهان و نورون لایه‌ی خروجی از وزن اتصال استفاده می‌شود. مقادیر این وزن‌ها به همراه مراکز و شعاع‌های نورون‌های لایه پنهان، در زمان آموزش مدل تعیین می‌شوند. نورون موجود در لایه خروجی با بهره‌گیری از رابطه ۱۱، مجموع وزن‌دار توابع گوسی لایه‌ی پنهان را محاسبه می‌کند.

$$y = \sum_{i=1}^{k} w_i \, h_i(X)$$ رابطه(۱۱)

ام لایه پنهان با نورون i وزن اتصال نورون w_i ، خروجی تنها نورون لایه خروجی شبکه است. y موجود در لایه‌ی خروجی است.

آموزش شبکه‌ی RBF در دو مرحله انجام می‌شود. ابتدا، مرکز و شعاع نورون‌های لایه‌ی میانی محاسبه می‌شوند. روش‌های مختلفی مانند انتخاب تصادفی، خود سازمان‌یافته و نظارت‌شده برای تعیین مرکز نورون‌های لایه میانی معرفی شده‌اند، که در این پژوهش به منظور دقت بیشتر از روش خود سازمان‌یافته ا ستفاده می شود. در مرحله‌ی دوم، وزن‌های بین لایه‌ی پنهان و لایه‌ی خروجی آموزش داده می‌شوند. روش مرسومی که برای این‌کار استفاده می‌شود بهره‌گیری از الگوریتم پس‌انتشار خطا ا ست که در دیگر شبکه‌ها مانند MLP ا ستفاده می شود. این الگوریتم، مقادیر وزن‌ها را به‌گونه‌ای تغییر می‌دهد که منجر به افزایش عملکرد مدل و یا به بیانی دیگر کاهش خطای مدل گردد.

۲-٤-٤. رگرسیون بردار پشتیبان (SVR)

رگرسیون بردار پشتیبان متداول‌ترین الگوریتم در میان الگوریتم‌های جداسازی خطی است که برای حل مسائل رگرسیون غیرخطی مناسب است. نحوه‌ی عملکرد مدل به شرح زیر می‌باشد: مجموعه داده M را در نظر بگیرید که شامل بردارهای ورودی x_i و خروجی متناظر y_i است. تعداد نمونه‌های این مجموعه داده برابر n است.

$$M = \left\{ \begin{matrix} (x_i, y_i) | \; i = 1,2, \dots, n \\ , x_i \in R^N \, , y_i \in R \end{matrix} \right\}$$ رابطه(۱۲)

در رگرسیون، به دنبال تخمین تابع $f(x_i)$ هستیم که خروجی‌های آن کم‌ترین فاصله را با مقادیر y_i داشته باشد. در رابطه ۱۳، δ خطای تصادفی با توزیع $N(0, \sigma^2)$ است.

$$y_i = f(x_i) + \delta$$ رابطه(۱۳)

به‌منظور حل یک مسئله رگرسیون غیرخطی با SVR، ابتدا با بهره‌گیری از رابطه‌ی ۱۴، ورودی‌ها به‌صورت غیرخطی به فضای ویژگی f با ابعاد زیاد که به‌صورت خطی با خروجی وابستگی دارند، نگاشت می شوند. در این رابطه، w بردار وزن، b مقدار بایاس و $\varphi(x)$ تابعی است که ورودی‌ها را از فضای R به فضای $R^{N \times h}$ تصویر می‌کند.

$$f(x_i) = w\varphi(x_i) + b \ \ |w \in R^{N\times h}, b \in R$$ رابطه(۱۴)

تابع خطی f به‌گونه‌ای اســـت که به اندازه ε از مقادیر واقعی انحراف دارد. برای داده‌هایی که خارج از باند باشند، از تابع جریمه‌ای استفاده می‌شود که با رابطه‌ی ۱۵ ارائه گردیده است.

$$L_\varepsilon\big(y_i, f(x_i)\big) = \begin{cases} 0 & |y_i - f(x_i)| \le \varepsilon \\ |y_i - f(x_i)| - \varepsilon & otherwise \end{cases}$$ رابطه(۱۵)

برای محاسبه‌ی ریسک عملیاتی تابع f از رابطه‌ی ۱۶ استفاده می‌شود.

$$R_{emp}[f] = \sum_1^n L_\varepsilon\big(y_i, f(x_i)\big)$$ رابطه(۱۶)

برای کاهش ریسک عملیاتی مربوط به داده‌ها با توجه به بهینه سازی تابع رگر سیون از رابطه ۱۷ بهره گرفته می‌شود.

$$J = \frac{1}{2}\|w\|^2 + CR_{emp}[f]$$ رابطه(۱۷)

برای داده‌هایی که مقدار $|y - f(x_i)|$ آن‌ها بیشـــتر از ε باشـــد، مقدار ξ_i^+ و یا ξ_i^- که مقدار تخطی را نشان می‌دهد، با رابطه‌های ۱۸و ۱۹ محاسبه می‌شود.

$$\xi_i^+ = y - f(x_i) - \varepsilon$$ رابطه(۱۸)

$$\xi_i^- = \ \varepsilon - y - f(x_i)$$ رابطه(۱۹)

همچنین، رابطه‌ی ۲۰ بین تابع جریمه و مقدارهای تخطی وجود دارد.

$$L_\varepsilon\big(y_i, f(x_i)\big) = \xi_i^+ + \xi_i^-$$ رابطه(۲۰)

برای تخمین تابع f از تابع هدف با رابطه‌ی ۲۱ استفاده می‌شود.

$$minimize \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^n(\xi_i^+ + \xi_i^-)$$ رابطه(۲۱)

$$\text{S.t. } \forall\, i$$
$$-y_i + f(x_i) + \varepsilon + \xi_i^+ \ge 0$$
$$y_i - f(x_i) + \varepsilon + \xi_i^- \ge 0$$
$$\xi_i^+, \xi_i^- \ge 0$$

سپس، از ضریب لاگرانژ برای ایجاد شکل دوگان رابطه ۲۱ ا ستفاده شده و ساده سازی انجام می شود. در صورتی که α_i^+ و α_i^-، به ترتیب ضریب قیدهای اول و دوم رابطه ۲۱ با شند، بعد از ساده‌سازی، رابطه ۲۲ حاصل می‌شود.

$$minimize \ \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n(\alpha_i^+ - \alpha_i^-)\big(\alpha_j^+ - \alpha_j^-\big) < \varphi(x_i)\cdot$$ رابطه(۲۲)

$$\varphi\big(x_j\big) > -\sum_{i=1}^n(\alpha_i^+ - \alpha_i^-)y_i + \varepsilon\sum_{i=1}^n(\alpha_i^+ + \alpha_i^-)$$
$$\text{S. t. } \begin{cases} \sum_{i=1}^n(\alpha_i^+ - \alpha_i^-) = 0 \\ \alpha.\alpha^* \in [0, C] \end{cases}$$

در مسـائل غیرخطی، ضـرب داخلی دو تابع $\varphi(x_i)$ و $\varphi\big(x_j\big)$ با تابع کرنل گوسـی، در رابطه ۲۳ جایگزین می‌شود.

$$K\big(x_i, x_j\big) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$ رابطه(۲۳)

درنهایت، تابع f با رابطه ۲۴ محاسبه می‌شود.

$$f(x) = \sum_{i=1}^n(\alpha_i^+ - \alpha_i^-)K(x_i, x) + b$$ رابطه(۲۴)

در این رابطه، محاسبه b با فرمول ۲۵ انجام می‌شود. SV بردار پشتیبان می‌باشد.

$$b = \frac{1}{n}\left\{\sum_{0<\alpha_i^-<C}\left[y_i - \sum_{x_j\in SV}(\alpha_j^+ - \alpha_j^-)K(x_i,x_j) - \varepsilon\right] + \sum_{0<\alpha_i^+<C}\left[y_i - \sum_{x_j\in SV}(\alpha_j^+ - \alpha_j^-)K(x_i,x_j) + \varepsilon\right]\right\} \quad \text{رابطه(۲۵)}$$

۳-٤-٤. پرسپترون چند لایه (MLP)

در این شبکه نیز مانندRBF، از سه لایه (یک لایه‌ی ورودی، یک لایه‌ی مخفی و یک لایه‌ی خروجی) استفاده می‌شود، با این تفاوت که بین لایه‌ی ورودی و لایه‌ی پنهان از وزن استفاده می شود. به‌طور معمول MLP به صورت یک مدل چند ورودی و تک خروجی می‌با شد. از لایه‌ی ورودی برای تزریق داده‌ها به شبکه ا ستفاده می شود. اگر مدل m ورودی دا شته با شد، لایه‌ی ورودی شامل m+1 نورون خواهد بود. اولین نورون، با مقدار ۱ به‌عنوان بایاس بوده و دیگر نورون ها، داده های ورودی را در یافت می‌کنند. دومین لا یه، با نام لا یه‌ی پنهان، مقادیر را از لایه‌ی ورودی دریافت کرده و بر روی آن‌ها یک عملیات محاسباتی انجام می‌دهد. این لا یه شامل n+1 نورون است. مشابه با اولین لایه، در اینجا هم اولین نورون بایاس است. دیگر نورون های این لا یه، یک تابع فعالیت را بر روی مجموع وزن‌دار خروجی لا یه‌ی قبل اعمال می‌کنند که خروجی آن با رابطه ۲۶ محاسبه می‌شود:

$$f_i = g\left(w1_{0,i} + \sum_{j=1}^{m} w1_{j,i}\, x_j\right) \quad \text{رابطه(۲۶)}$$

در این رابطه، $w1_{0,i}$ وزن اتصال بایاس با iامین نورون لایه‌ی پنهان و $w1_{ji}$ وزن اتصال بین jامین نورون لایه‌ی ورودی و iامین نورون در لایه‌ی پنهان است. همچنین، x_j ورودی jام به شبکه عصبی است. مشابه عملیاتی که درلایه‌ی پنهان انجام می‌شود، نورون‌های لایه‌ی خروجی تابع فعالیتی را بر روی مجموع وزن‌دار خروجی لا یه‌ی پیش از خود اعمال کرده و مقدار خروجی با رابطه‌ی ۲۷ محاسبه می‌شود:

$$y_i = h\left(w2_{0,i} + \sum_{j=1}^{n} w2_{j,i}\, f_j\right) \quad \text{رابطه(۲۷)}$$

$w2_{0,i}$ وزن اتصال بایاس با iامین نورون لایه‌ی خروجی و $w2_{ji}$ وزن اتصال بین jامین نورون لایه‌ی پنهان و iامین نورون در لایه‌ی خروجی است. همچنین، f_j مقداری خروجی نورون jام از لایه‌ی پنهان است. لازم به ذکر است که در اغلب مسائل رگرسیون، از یک نورون در لایه‌ی خروجی استفاده می‌شود.

٤-٤-٤. شبکه عصبی عمیق (DNN)

در سال‌های اخیر شبکه عصبی عمیق توجه جامعه علمی را به خود جلب کرده است. هدف اصلی آن، مدل‌سازی داده‌های پیچیده دنیای واقعی به‌وسیله استخراج ویژگی‌های مناسب از داده‌های مرتبط است[۱۸]. یکی از چالش‌های اصلی این شبکه، فقدان یک نظریه ساختاری مناسب برای ایجاد آن است. به بیانی دیگر، تئوری خاصی برای چیدمان عناصر شبکه وجود ندارد.

بنابراین، برای انتخاب ساختار مناسب لازم است که مدل به‌صورت تجربی مورد مطالعه قرار گیرد. برای این منظور از تکنیک هایی مانند تغییر تعداد لایه ها، تعداد نورون ها و آزمایش توابع فعال سازی مختلف، استفاده نموده و ساختارهای مختلف این شبکه را مورد ارزیابی قرار می‌دهیم تا در نهایت ساختار مناسب انتخاب شود.

مدل پیشنهادی در این مقاله شامل یک لایه ورودی، سه لایه پنهان و یک لایه خروجی می‌باشد. در سه لایه‌ی پنهان به ترتیب از ۲۰، ۱۰ و ۵ نورون استفاده شده است. از تابع Dense موجود در کتابخانه Keras ازکتابخانه‌های موجود در مجموعه tensorflow در زبان پایتون برای ایجاد لایه‌های شبکه DNN استفاده شده است. برای نورون‌های لایه پنهان از تابع فعال‌سازی ReLU بهره گرفته شده است که یکی از پرکاربردترین و ساده‌ترین توابع فعال سازی غیرخطی است. دلیل اصلی استفاده از این تابع این است که با تعداد زیادی ورودی اشباع نمی‌شود. همچنین خطای مدل به سادگی در همه لایه‌ها منتشر شده که باعث آموزش بهتر مدل می‌گردد. مشکل تابع ReLU این است که فقط برای مقادیر بزرگتر از صفر مقدار می‌گیرد. به همین دلیل، در لایه خروجی که حاوی تنها نورونی است که مقدار شاخص پیش‌بینی شده را ارائه می‌کند از تابع فعالیت Sigmoid استفاده می‌شود. یکی دیگر از تنظیمات صورت گرفته برای طراحی مدل انتخاب مقدار دوره[1] و اندازه دسته[2] است. در هر دوره، از تمام داده‌های آموزش برای به‌روزرسانی وزن‌های شبکه استفاده می‌شود. اندازه دسته نیز تعداد نمونه‌های آموزشی را مشخص می‌کند که به طور هم‌زمان در یک دوره، به شبکه وارد شده و متغیر وابسته یعنی بازده روزانه شاخص بورس، پیش‌بینی می‌شود. با انجام آزمایشات مختلف و مقایسه نتایج حاصل، مقدار دوره و اندازه دسته به ترتیب ۱۰۰ و ۵ برای مدل تعیین می‌شود. همچنین، برای محاسبه نرخ یادگیری تطبیقی برای هر پارامتر و به‌روزرسانی وزن‌های لایه پنهان از بهینه‌ساز Adam استفاده می‌شود. بهینه‌ساز Adam یک روش تصادفی گرادیان نزولی[3] است که بر اساس تخمین انطباقی لحظه‌های مرتبه اول و مرتبه دوم است. Adam، مناسب برای آموزش مدل‌های یادگیری عمیق است. پارامترهایی که با استفاده از این بهینه‌ساز یادگرفته می‌شوند عبارتند از: نرخ یادگیری[4] یا اندازه گام[5]، نرخ تجزیه نمایی برای برآورد لحظه اول، نرخ تجزیه نمایی برای برآورد لحظه دوم و اپسیلون که عدد بسیار کوچکی است و برای جلوگیری از تقسیم بر صفر در پیاده‌سازی استفاده می‌شود.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Epoch
[2] Batchsize
[3] Stochastic Gradient Descent
[4] learning Rate
[5] Step Size

معیارهای ارزیابی مدل‌های پیش‌بینی

به‌منظور ارزیابی مدل در مرحله‌ی پیاده‌سازی، روش‌های متنوعی وجود دارد. هرچه حجم نمونه برای داده‌های آموز شی غنی‌تر (تعداد م شاهدات کافی) با شد، امکان ر سیدن به نتایج مطلوب‌تر برای داده‌های آزمون محتمل‌تر است. بدین منظور در بیشتر مطالعات بخش عمده‌ای از مشاهدات را به داده‌های آموزش اختصاص داده‌اند[۳۷]. این مقاله، از روش اعتبار سنجی متقابل K-fold[1]، برای انتخاب مجموعه های آموزش و آزمایش بهره برده اســت. از مزایای اســتفاده از روش اعتبارســنجی متقابل K-fold این اســت که این روش هم برای مجموعه داده‌های بزرگ و هم برای مجموعه داده‌های کوچک مناسب است. همچنین در این روش، از همه نمونه‌های مجموعه داده برای آزمایش مدل اســتفاده می‌شــود. بنابراین یک معیار ارزیابی کامل و منصــفانه اســت. درصــورتی‌که در روش‌هایی که داده‌ها به دو دسته‌ی آموزش و آزمایش (به عنوان مثال ۲۰–۸۰، ۸۵–۱۵، ۷۰–۳۰ و...) تقسیم می شوند، با هر بار اجرای مدل نتیجه‌ی متفاوتی می‌تواند ارائه شود. زیرا دقت مدل به شدت وابسته به داده‌های آموزش و آزمایش بوده و این درحالی است که انتخاب این دو دسته به صورت تصادفی انجام می‌شود.

در این رویکرد K مرحله‌ای مجموعه داده به ۱۰دســته[2] متمایز تقســیم‌بندی می‌شــود. در هر مرحله، ۱–۱۰ دســته برای آموزش مدل و یک دســته برای آزمایش مدل انتخاب می‌شــوند. به‌طوری‌که در هر مرحله ۹۰ درصــد داده‌ها برای مدل‌ســازی و ۱۰ درصــد برای آزمون اســتفاده می‌شــوند. لازم به ذکر اســت که انتخاب مقدار ۱۰ با توجه به متداول بودن اســتفاده از این عدد برای اعتبار سنجی م سائل رگر سیون تو سط محققین می‌با شد[٦]. البته، پیش از تق سیم‌بندی داده‌ها، ابتدا جایگشــتی بر روی مجموعه‌داده انجام شــده تا نمونه‌ها به‌هم‌ریخته شــوند. این به‌هم‌ریختگی باعث تنوع نمونه‌ها در توزیع داده‌ها می شود. شایان ذکر است در این تحقیق، یکی از پیش‌پردازش‌های اولیه بر روی مجموعه داده‌های مورد بررســی، افزودن ویژگی‌های آماری به آن اســت. ویژگی‌های آماری برای هر نمونه، شــامل مقادیر پیشــین هر ویژگی و میزان تغییرات آن‌ها می‌باشــد. با این رویکرد، با بهره‌گیری از ویژگی‌های آماری، اســتقلال هر نمونه نســبت به دیگر نمونه‌ها فراهم شــده اســت. بر این اســاس، برای پیش‌بینی شــاخص بورس، می‌توان از مدل‌هایی غیر از مدل‌های سری زمانی استفاده کرد. این روش می‌تواند علاوه بر کمک به انتخاب ویژگی‌های مناسب، باعث افزایش دقت پیش‌بینی گردد.

در این پژوهش به‌منظور بررســی دقت مدل‌های پیش‌بینی، از معیارهای میانگین قدر مطلق خطا[3] (MAE)، میانگین مربع خطا[4] (MSE)، و ریشه میانگین مربع خطا[5] (RMSE) استفاده شده

[1] K-fold Cross Validation
[2] 10-fold Cross Validation
[3] Mean Absolute Error
[4] Mean Square Error
[5] Root Mean Square Error

است که با رابطه‌ی ۲۸، ۲۹ و ۳۰ محاسبه می‌شوند.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|t_i - y_i|$$ رابطه(۲۸)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(t_i - y_i)^2$$ رابطه(۲۹)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t_i - y_i)^2}$$ رابطه(۳۰)

N تعداد نمونه‌ها، t_i مقدار هدف و y_i مقدار پیش‌بینی‌شده برای نمونه‌ی iام است.

٤- تحلیل داده‌ها و یافته‌ها

اولویت‌بندی ویژگی‌های مؤثر بر شاخص با استفاده از روابط تخمین mRMR

روش‌های آماری مورد ا ستفاده برای تخمین mRMR در این مقاله، MID و FCD می‌با شند. این روش‌ها با آن‌که تخمین قابل‌قبولی از mRMR دارند، ولی پیچیدگی محاسباتی آن‌ها پایین است و از سرعت و دقت قابل قبولی برخوردار هستند. بنابراین ویژگی‌های موجود در مجموعه داده ورودی را یک‌بار با استفاده از الگوریتم MID و بار دیگر با استفاده از FCD، در بردار ζ بر اساس میزان وابستگی ویژگی‌ها با مقدار خروجی اولویت‌بندی می‌شوند. سپس در $|\zeta|$ گام، زیرمجموعه‌هایی شامل $1 \leq j \leq |\zeta|$ عنصر ابتدایی از ζ را در مدل پیشنهادی ارزیابی کرده و میزان خطای پیش‌بینی برای هر یک اندازه‌گیری می‌گردد. ترتیب اولویت ویژگی‌ها با بهره‌گیری از MID و FCD به‌ترتیب در جدول‌های ۲ و ۳ نمایش داده شده است.

جدول۲ . اولویت‌بندی ویژگی‌ها با الگوریتم MID با توجه به شاخص بورس یک روز آینده

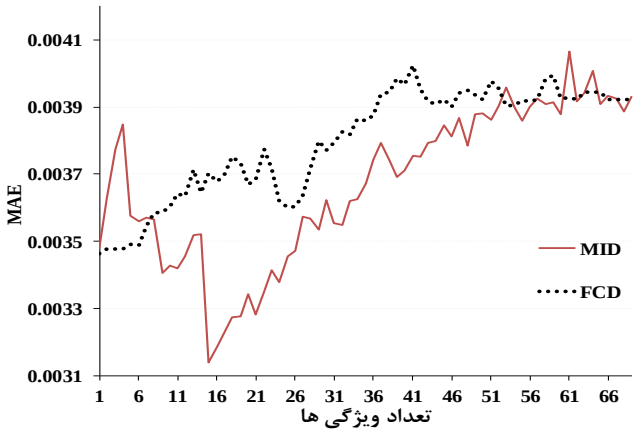
شماره	نام ویژگی	شماره	نام ویژگی	شماره	نام ویژگی	شماره	نام ویژگی	شماره	نام ویژگی
۱	RDP5	۱۵	IC502	۲۹	U3	۴۳	InI2	۵۷	GPd4
۲	TV1	۱۶	OP1	۳۰	OPd4	۴۴	MPP30	۵۸	P/Ed2
۳	P/Ed3	۱۷	PIC501	۳۱	P/E2	۴۵	Dd3	۵۹	Ud2
۴	D1	۱۸	U1	۳۲	PIC503	۴۶	TEDPIX 3	۶۰	Dd1
۵	LCI301	۱۹	FI3	۳۳	U2	۴۷	EMA20	۶۱	OPd1
۶	RDP10	۲۰	OP3	۳۴	MPP120	۴۸	TEDPIX 2	۶۲	GPd3
۷	GP2	۲۱	P/E1	۳۵	InI1	۴۹	OPd3	۶۳	Dd4
۸	OP2	۲۲	D3	۳۶	InI3	۵۰	P/Ed4	۶۴	EMA10
۹	LCI303	۲۳	RDP20	۳۷	LCI302	۵۱	GPd1	۶۵	MA30
۱۰	GP3	۲۴	FI2	۳۸	GPd2	۵۲	Ud1	۶۶	MA5
۱۱	FI1	۲۵	P/E3	۳۹	PIC502	۵۳	Dd2	۶۷	MA10
۱۲	GP1	۲۶	TV3	۴۰	IC501	۵۴	P/Ed1	۶۸	EMA50
۱۳	RDP15	۲۷	D2	۴۱	Ud4	۵۵	Ud3	۶۹	EMA200
۱۴	TV2	۲۸	TEDPIX 1	۴۲	IC503	۵۶	OPd2		

جدول۳. اولویت‌بندی ویژگی‌ها با الگوریتم FCD با توجه به شاخص بورس یک روز آینده

شماره	نام ویژگی	شماره	نام ویژگی	شماره	نام ویژگی	شماره	نام ویژگی	شماره	نام ویژگی
۱	RDP5	۱۵	GPd2	۲۹	TEDPIX3	۴۳	MA30	۵۷	D1
۲	RDP10	۱۶	GPd3	۳۰	LCI302	۴۴	FI3	۵۸	P/Ed4
۳	IC501	۱۷	Dd3	۳۱	IC502	۴۵	P/E2	۵۹	U1
۴	InI1	۱۸	Dd4	۳۲	InI2	۴۶	GP3	۶۰	P/Ed3
۵	PIC501	۱۹	Ud1	۳۳	TEDPIX2	۴۷	GP2	۶۱	P/Ed۳
۶	LCI301	۲۰	GPd4	۳۴	EMA200	۴۸	FI2	۶۲	GP1
۷	RDP15	۲۱	Ud2	۳۵	U2	۴۹	D2	۶۳	OP3
۸	RDP20	۲۲	Ud3	۳۶	P/E3	۵۰	OP1	۶۴	OP2
۹	P/E1	۲۳	PIC503	۳۷	MPP120	۵۱	OPd2	۶۵	TV2
۱۰	FI1	۲۴	Ud4	۳۸	MA5	۵۲	OPd3	۶۶	P/Ed1
۱۱	Dd1	۲۵	PIC502	۳۹	EMA10	۵۳	OPd4	۶۷	TV3
۱۲	MPP30	۲۶	LCI303	۴۰	MA10	۵۴	U3	۶۸	TV1
۱۳	Dd2	۲۷	IC503	۴۱	EMA20	۵۵	OPd1	۶۹	TEDPIX 1
۱۴	GPd1	۲۸	InI3	۴۲	EMA50	۵۶	D3		

لازم به ذکر است که در این بخش از بررسی‌ها، از مدل RBF استفاده می‌شود. مقدار MAE در شـــکل ۲ برای پیش‌بینی شـــاخص بورس یک روز آینده، برای هر زیرمجموعه از ویژگی‌های انتخاب شده نمایش داده شده است. با استفاده از آن می‌توان مقایسه‌ای بین بهره‌گیری از MID و FCD در اولویت‌بندی ویژگی‌ها و اثر افزایش تعداد ویژگی‌های مدل پیش‌بینی دا شت. یکی از نتایج قابل مشاهده در شکل این است که با افزایش تعداد ویژگی‌های اولویت‌بندی شده با FCD خطای مدل افزایش می‌یابد. در صـــورتی‌که با بهره‌گیری از اولویت‌بندی با MID، افزایش تعداد ویژگی‌ها ابتدا باعث افزایش دقت مدل می‌گردد ولی از ویژگی ۱۵ به بعد با کاهش دقت مدل روبه‌رو ه ستیم. با مقای سه نتایج حا صل از MID و FCD، می‌توان گفت MID نتیجه بهتری در اولویت‌بندی ویژگی‌ها ارائه نموده اســـت. براین‌اســـاس، الگوریتم MID برای اولویت‌بندی ویژگی‌ها انتخاب گردید.

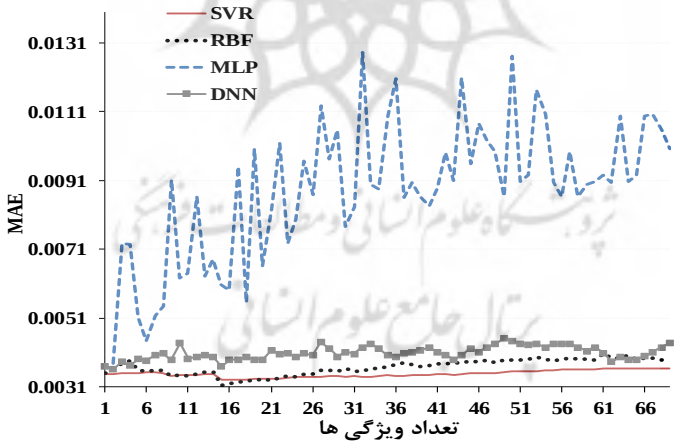
پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



شکل۲. استفاده از مدل RBF و ویژگی‌های اولویت‌بندی شده با روش‌های MID و FCD برای پیش‌بینی روزانه شاخص بورس

انتخاب مدل مناسب برای پیش‌بینی شاخص بورس

در این بخش جهت انتخاب مدل مناسب برای پیش‌بینی شاخص بورس به مقایسه چهار مدل پیش‌بینی به‌نام‌های RBF، SVR، MLP وDNN با ویژگی‌های اولویت‌بندی شده با MID، پرداخته می شود. نتایج استفاده از این روش برای چهار مدل در شکل ۳ نمایش داده شده است. همان‌طور که مشاهده می‌شود، با توجه به این‌که مدل‌های RBF SVR و DNN عملکرد بهتری نسبت به مدل MLP ارائه داده‌اند، تصمیم بر حذف مدل MLP گردید. همچنین برای مقایسه بهتر نتایج استفاده از این سه مدل بعد از حذف MLP، نتیجه مربوط به آن‌ها مجدد در شکل ۴ نمایش داده می‌شود.

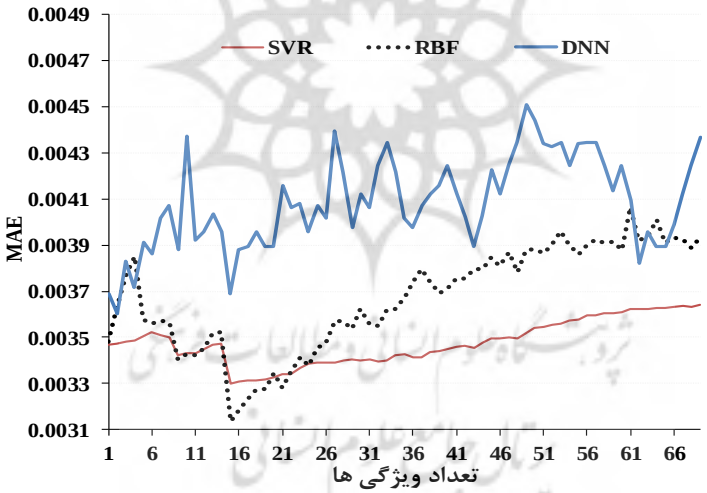


شکل ۳. مقایسه استفاده از مدل‌های RBF، SVR، MLP و DNN در پیش‌بینی روزانه شاخص بورس

در این حالت، بهتر می‌توان این سه مدل را با یک‌دیگر مقایسه کرد و اثر افزایش تعداد ویژگی‌ها در دقت مدل‌های پیش‌بینی شاخص بورس را مشخص نمود. با بهره‌گیری از شکل ۴ می‌توان به نکاتی اشاره کرد. در مدل DNN افزایش تعداد ویژگی‌ها منجر به بهبود مدل نشده است. کمترین خطا در استفاده از مدل RBF و SVR، با ۱۵ ویژگی حاصل شده است. به‌طوری‌که در این دو مدل، افزایش تعداد ویژگی‌ها تا ۱۵ ویژگی باعث بهبود دقت پیش‌بینی شاخص می‌شود. بعد از آن با کاهش دقت مواجه هستیم. از دیگر نکات قابل ملاحظه این است که اثر ۲ ویژگی ۹ و ۱۵ یعنی بازده شاخص ۳۰ شرکت بزرگ در۳ روز گذشته و بازده شاخص ۵۰ شرکت فعال‌بورس در ۲ روز گذشته، تأثیر قابل توجهی در بهبود عملکرد هر دو مدل داشته‌اند. یکی دیگر از نکات حائز اهمیت، تغییر تعداد ویژگی‌ها در مدل SVR می‌باشد که نوسانات کمی در مقدار خطای مدل ایجاد می‌کند. با این وجود اگر بخواهیم مدلی مناسب با بهترین دقت و کمترین تعداد ویژگی انتخاب کنیم مدل RBF با ۱۵ویژگی دارای خطای کمتری می‌باشد. نتایج در جدول ۴ ارائه شده است.

جدول ۴. اثر دو ویژگی ۹ و ۱۵ بر عملکرد مدل‌های SVR و RBF

شماره ویژگی	نام مدل	MAE	MSE	RMSE
شماره ۹ (LCI303)	RBF	۰/۰۰۳۴۰۶	۰/۰۰۰۰۲۹	۰/۰۰۵۴۶۳
	SVR	۰/۰۰۳۴۲۲	۰/۰۰۰۰۳۰	۰/۰۰۵۴۸۷
شماره ۱۵ (IC502)	RBF	۰/۰۰۳۱۳۸	۰/۰۰۰۰۲۴	۰/۰۰۴۹۵۵
	SVR	۰/۰۰۳۳۰۱	۰/۰۰۰۰۲۹	۰/۰۰۵۳۳۹



شکل۴. مقایسه استفاده از مدل‌های SVR، RBF و DNN در پیش‌بینی شاخص بورس

الگوریتم پیشنهادی جهت انتخاب ویژگی‌های مناسب برای مدل پیش‌بینی شاخص بورس

در الگوریتم پیشنهادی برای انتخاب ویژگی‌های مناسب، تعدادی از ویژگی‌ها توسط محقق انتخاب می‌شوند. اگر الگوی تغییر مقدار خطا با افزایش تعداد ویژگی‌ها که در شکل۴ نمایش داده شده است با دقت بیشتری بررسی کنیم، نکاتی در آن دیده می‌شود. در مدل SVR، دو ویژگی بازده شاخص ۳۰ شرکت بزرگ در۳ روز گذشته ($LCI30_3$) و بازده شاخص ۵۰ شرکت فعال‌بورس در ۲ روز گذشته ($IC50_2$)، باعث کاهش قابل توجه خطای پیش‌بینی شده‌اند. در مدل RBF، علاوه بر دو ویژگی فوق، ویژگی شماره ۵ یعنی بازده شاخص ۳۰ شرکت بزرگ در ۱روز گذشته ($LCI30_1$) نیز سبب کاهش زیادی در مقدار خطا شده است. بر این اساس، تصمیم بر آن شد که برای اولویت‌بندی ویژگی‌ها ابتدا این سه ویژگی را به‌عنوان ویژگی‌های معرفی شده، انتخاب کرده و سپس به اولویت‌بندی سایر ویژگی‌ها با الگوریتم MID پرداخته شود. ویژگی‌ها با اولویت جدیدی که برای آن‌ها به‌دست آمده به مدل RBF به منظور پیش‌بینی شاخص بورس به‌ترتیب داده می‌شوند. الگوریتم پیشنهادی به شرح زیر می‌باشد:

$FS = \{\text{Feature 15, Feature 9. Feature 5}\}$
$SPE = \text{StockPredictionError}(FS)$
$\zeta = \text{mRMR} - \text{MID}(\text{AllFeatures}, FS)$
$idx = 1$
while $idx \leq |\zeta|$
$FS = FS \cup \zeta(idx)$
$SPE = SPE \cup \text{StockPredictionError}(FS)$
$idx = idx + 1$
end
Plot(SPE)

FS مجموعه ویژگی‌های انتخابی است. با توجه به نتایج حاصل از شکل ۴، ابتدا ویژگی‌های ۱۵، ۹ و ۵ را به مجموعه FS وارد می‌کنیم. خطای حاصل از استفاده از این ویژگی‌ها برای پیش‌بینی شاخص بورس با استفاده از تابع StockPredictionError محاسبه می‌گردد. مقدار به دست آمده به مجموعه SPE اضافه می‌شود. سپس، اولویت‌بندی ویژگی‌ها با MID صورت می‌گیرد. ورودی MID، همه‌ی ویژگی‌ها به همراه مجموعه FS است. MID سایر ویژگی‌ها را با توجه به ویژگی‌های اولیه انتخابی FS اولویت‌بندی می‌کند و به مجموعه ζ نسبت می‌دهد. در ادامه از یک حلقه تکرار استفاده شده است. در هر تکرار، یکی از ویژگی‌های مجموعه ζ به FS افزوده شده و مجدد خطای حاصل از پیش‌بینی شاخص بورس با مجموعه به‌روز شده FS محاسبه و به مجموعه SPE اضافه می‌گردد. در نهایت با استفاده از تابع plot نمودار خطای حاصل از افزایش تعداد ویژگی‌ها در پیش‌بینی شاخص بورس ترسیم می‌شود.

شـایان ذکر اسـت که در این ارزیابی نیز از روش اعتبارسـنجی K-fold (K=10) اسـتفاده می شود. همچنین نتیجه این محا سبات که اثر افزایش تعداد ویژگی‌ها با اولویت م شخص شده را در دقت مدل RBF نشان می‌دهد، در شکل ۵ نمایش داده شده است. در این شکل، دو حالت با هم مقایسـه شـده‌اند. در حالت اول، پس از اولویت‌بندی ویژگی‌ها با الگوریتم MID، اثر افزایش تعداد ویژگی‌ها در دقت مدل پیش‌بینی شاخص بورس ن شان داده شده ا ست. در حالت دوم، اثر افزایش تعداد ویژگی‌ها در دقت مدل پیش‌بینی شـاخص بورس پس از اولویت‌بندی ویژگی‌ها با ا ستفاده از الگوریتم پی شنهادی تحت عنوان ISF_MID[1] نمایش داده شده ا ست. با مقای سه‌ی نمودارهای موجود در شکل ۵ می‌توان به این نتیجه ر سید که، مدل RBF با ویژگی‌های انتخابی توسـط الگوریتم پیشـنهادی ISF_MID، تنها با ۷ ویژگی، بهترین نتیجه و دقت را ارائه کرده اسـت. در صـورتی‌که مدل RBF با ویژگی‌های انتخابی توسـط الگوریتم MID، با ۱۵ ویژگی بیشترین دقت را دارد. نتایج در جدول ۵ نشان داده شده است.

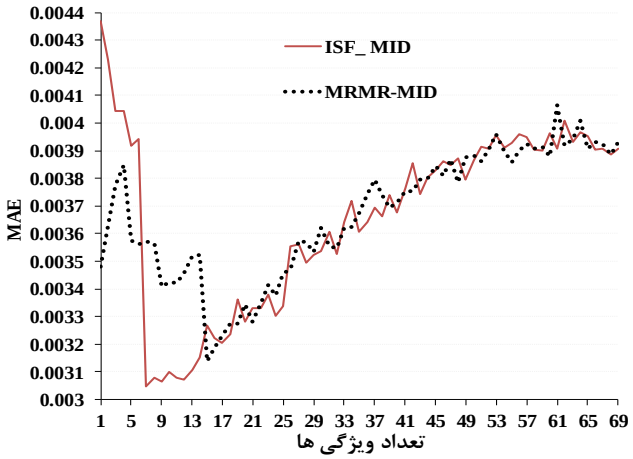
جدول ۵. نتایج استفاده از الگوریتم ISF_MID و الگوریتم MID در انتخاب ورودی‌های مدل RBF

	ISF_MID (الگوریتم پیشنهادی)	**MID**
تعداد ویژگی	۷	۱۵
MAE	۰/۰۰۳۰۵	۰/۰۰۳۱۴
MSE	۰/۰۰۰۰۲۳	۰/۰۰۰۰۲۴
RMSE	۰/۰۰۴۸۱۹	۰/۰۰۴۹۶

با توجه به نتایج ارائه شده در جدول ۵ می‌توان به این نتیجه ر سید که مدل RBF با الگوریتم ISF_MID، در مقایسـه با مدل RBF با الگوریتم MID، عملکرد بهتری دارد. به‌طوری‌که مقدار MAE با ۷ ویژگی ابتدایی، برابر با ۰/۰۰۳ است. این ویژگی‌ها به ترتیب عبارتند از: بازده شاخص ۵۰ شرکت فعال‌بورس در ۲ روز گذ شته، بازده شاخص ۳۰ شرکت بزرگ در۳ روز گذ شته، بازده شاخص ۳۰ شرکت بزرگ در۱روز گذ شته، تغییر ن سبی حجم معاملات در ۳ روز گذ شته، تفاوت نسبی درصد بازده شاخص در ۱۰روز گذشته، قیمت دلار در ۳ روز گذشته، و تفاوت نسبی درصد بازده شاخص در ۵ روز گذشته.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Initial Selected Features_ Mutual Information Difference



شکل ۵. مقایسه نتیجه استفاده از الگوریتم ISF_MID با الگوریتم MID در انتخاب ورودی‌های مدل RBF برای پیش‌بینی روزانه شاخص کل بورس

۵. بحث و نتیجه‌گیری

در این پژوهش سعی شده است با رویکردی کاملا جدید و منحصر به‌فرد، مدلی برای پیش‌بینی روزانه شاخص کل بورس اوراق بهادار تهران ارائه شود. در این راستا به‌منظور انتخاب ویژگی‌های مناسب برای ورودی مدل، الگوریتمی پیشنهاد شده است که باعث افزایش دقت مدل ارائه شده گردید. به طور خلاصه؛ اقداماتی که در این پژوهش صورت گرفت که باعث تمایز آن با سایر پژوهش‌ها در این حوزه می‌شود، به شرح زیر می‌باشند:

– جمع‌آوری یک مجموعه کامل از ویژگی‌های تأثیرگذار بر شاخص بورس براساس مبانی تئوریک و بررسی تحقیقات پیشین؛ به‌طوری‌که ۶۹ ویژگی از ۱۶ شاخص مورد بررسی قرار گرفت که هر دو رویکرد تکنیکی و بنیادی را شامل می‌شوند. شایان ذکر است که تجزیه و تحلیل فنی و بنیادی دو رویکرد اصلی برای تحلیل بازارهای مالی به‌حساب می‌آیند[۳۸]؛ و برای تصمیم‌گیری در بازار مالی و سرمایه‌گذاری در بازار سهام و کسب سود با ریسک کم، باید دو رویکرد گفته شده را همزمان به‌کار گرفت[۱].

– این مقاله، از روش اعتبارسنجی متقابل K-fold، برای انتخاب مجموعه‌های آموزش و آزمایش بهره برده‌است. این روش باعث می شود همه داده‌ها یک‌بار برای آموزش و یک‌بار برای آزمایش به‌کار روند، که این امر منجر به افزایش دقت و مفید واقع شدن مدل مورد نظر در عمل می‌گردد. در پژوهش‌های مدل سازی و پیش‌بینی متداول است که مقدار ۵ یا ۱۰ خوشه را برای K در نظر می‌گیرند. در این پژوهش از روش اعتبارسنجی ۱۰ خوشه‌ای به‌منظور افزایش قابلیت تعمیم مجموعه داده، استفاده شده است. به‌طوری‌که مجموعه داده به ۱۰دسته متمایز تقسیم‌بندی می‌شود. در هر مرحله، ۱–۱۰ دسته برای آموزش مدل و یک دسته برای آزمایش مدل انتخاب

می‌شوند. در صورتی‌که پیش‌تر در پژوهش‌های صورت گرفته در این حوزه، بخشی ازداده‌ها برای آموزش و بخشی دیگر برای آزمایش مورد استفاده واقع می‌شدند.

– برخلاف پژوهش‌های صورت گرفته در این حوزه که از مقدار همبستگی ویژگی‌های ورودی نسبت به خروجی مدل برای انتخاب ویژگی‌های مؤثر استفاده می‌شود، در این پژوهش برای انتخاب ویژگی‌های مناسب اقدام به اولویت‌بندی ویژگی‌ها با استفاده از معیار mRMR، گردید. این روش آماری؛ ویژگی‌های مؤثر را براساس وابستگی بین داده‌ها انتخاب می‌کند. به‌طوری‌که علاوه‌بر توجه به همبستگی هر ویژگی با خروجی مدل، همبستگی بین ویژگی‌های ورودی با یکدیگر نیز بررسی می شود و ویژگی‌ها با توجه به بیشینه سازی معیار وابستگی آماری مجموعه ویژگی‌ها با ویژگی هدف، و کمینه کردن اطلاعات متقابل در بین مجموعه ویژگی‌های انتخابی، گزینش می‌شوند. در این راستا برای تخمین mRMR، دو روش MID و FCD مورد بررسی قرار گرفتند و با مقایسه نتایج این دو، MID که نتیجه بهتری برای تخمین mRMR ارائه نمود، برای اولویت‌بندی ویژگی‌ها مورد استفاده قرار گرفت.

– به‌منظور انتخاب مدل مناسب برای پیش‌بینی شاخص بورس، ۴ مدلMLP، SVR، RBF و DNN که از مهم‌ترین و بدیع‌ترین مدل‌های پیش‌بینی می‌باشند و دقت آن‌ها در پژوهش‌های پیشین تأیید شده است، مورد مقایسه قرار گرفتند. شایان ذکر است که مدل DNN در سال‌های اخیر توجه محققان را برای پیش‌بینی شاخص بورس، به خود جلب نموده است و تاجایی‌که اطلاع داریم، این اولین تلاش برای بررسی تأثیر انتخاب ویژگی در مدل شبکه عصبی عمیق برای پیش‌بینی شاخص بورس در ایران می‌باشد.

– در بررسی مدل‌های مورد استفاده علاوه‌بر مقایسه آن‌ها به‌صورت کلی، اقدام به مقایسه مدل‌ها براساس تعداد ویژگی‌هایی که منجر به کاهش خطا و در نتیجه افزایش دقت پیش‌بینی می شوند، مبادرت شد. سپس به انتخاب مدل‌های مناسب برای پیش‌بینی اقدام گردید.

– الگوریتمی تحت عنوان ISF_MID به‌منظور انتخاب ویژگی‌های مناسب برای مدل پیش‌بینی روزانه شاخص بورس اوراق بهادار تهران پیشنهاد گردید که با توجه به این الگوریتم می‌توان با تعداد معدودی ویژگی (۷ ویژگی) به پیش‌بینی با بیشترین دقت اقدام نمود.

با توجه به نتایج این مطالعه، پیشنهاد می‌شود در پژوهش‌های آتی به‌منظور بررسی عملکرد الگوریتم ISF_MID، مجموعه داده‌های مورد بررسی گسترش یابد و/ یا از دیگر تکنیک‌های هوشمند به‌عنوان مدل پیش‌بینی استفاده شود. همچنین برای تکمیل آن می‌توان از داده‌های بورس دیگر کشورها استفاده نمود و عملکرد الگوریتم پیشنهادی را در این بازارها مورد مطالعه قرار داد.

در پایان خاطر نشان می‌کنیم که از نتایج این مطالعه پژوهشگران حوزه مدل‌سازی و سرمایه‌گذاران اعم از حقیقی و حقوقی می‌توانند استفاده کنند.

منابع

1. Arévalo, R., García, J., Guijarro, F., & Peris, A. (2017). A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting . *Expert Systems with Applications*, 81, 177-192.

2. Asadi, Sh., Hadavandi, E., Mehmanpazir, F., & Nakhostin, M. M. (2012). Hybridization of evolutionary Levenberg-Marquardt neural networks and data pre-processing for stock market prediction. *Knowledge-Based Systems, 35*, 245-258.

3. Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques -Part II: Soft computing methods. *Expert Systems with Applications*, 36(3), 5941-5950.

4. Badiei, H., Rezazadeh, R., & Mahmoudi, H. (2017). Forecasting Stock Price Trend by Artificial Neural Networks (Case Study: Isfahan Oil Refinery Company). *Financial Engineering and Portfolio Management*, 8(31), 167-185. (in Persian)

5. Bajalan, S., Fallahpour, S., & Dana, N. (2016). The Stock Trend Prediction Using Volume Weighted Support Vector Machine with a Hybrid Feature Selection Method to Predict the Stock Price Trend in Tehran Stock Exchange. *Financial Management Strategy*, 4(3), 121-148. (in Persian)

6. Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*, 1, 542-545.

7. Botshekan, M., & Mohseni, H. (2018). Investigation volatility spillovers between oil market and stock index return. *Journal of Investment Knowledge*, 7(25), 267-284. (in Persian)

8. Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. F., Nobrega, J. P., & Oliveira A. L. I. (2016). Computational Intelligence and Financial Markets: A Survey and Future Directions, *Expert Systems With Applications*, 55, 194-211.

9. Chong, E., Han, Ch., & Park F. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies, *Expert Systems With Applications*, 83,187-205.

10. Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(2), 185-205.

11. Enke, D., Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, 29(4), 927-940.

12. Faghihi Nezhad, M., & Minaei, B. (2018). Prediction of Stock Market Behavior Based on Artificial Neural Networks through Intelligent Ensemble Learning Approach. *Industrial Management Journal*, 10(2), 315-334. (in Persian)

13. 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)

14. Grigoryan, H. (2016). Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA). *Database Systems Journal*, 7(1), 12-21.

15. Gunduz, H., & Cataltepe, Z. (2015). Borsa Istanbul (BIST) daily prediction using financial news and balanced feature selection. *Expert Systems with Applications*, 42 (22), 9001-9011.

16. Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert systems with application*, 38, 10389-10397.

17. Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124, 226-251.

18. Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*(5786), 504-507.

19. Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, Kp. (2018). NSE stock market prediction using deep-learning models. *Procedia Computer Science*, 132: 1351-62.

20. Hu, Y., Liu, K., Zhang, X., Su, L., Ngai, E. W. T., & Liu, M. (2015). Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review. *Applied Soft Computing*, 36, 534-51.

21. Huang, C. F. (2012). A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing*, 12(2), 807-818.

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

23. Lam, M. (2004). Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis. *Decision Support Systems*, 37(4), 567-581.

24. Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning*, (pp. 609-616). Montreal, Canada.

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

26. Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, 164, 163-173.

27. Mandal, M., & Mukhopadhyay, A. (2013). An improved minimum redundancy maximum relevance approach for feature selection in gene expression data. Procedia Technol,10, 20-27.

28. Monfared, J., & Alinejad, M., & Metghalchi, S. (2012). A Comparative Study of Neural Network Models With Box Jenkins Methodologies in Prediction of Tehran Price Index(TEPIX). *Financial Engineering and Securities management (Portfolio Management)*, 3(11), 1-16. (in Persian)

29. Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment Analysis on Social Media for Stock Movement Prediction. *Expert Systems with Applications*, 42(24), 9603-9611.

30. Ou, P., & Wang, H. (2009). Prediction of stock market index movement by ten data mining techniques. *Modern Applied Science*, 3(12), 28-42.

31. Oussar, Y., & Dreyfus, G. (2000). Initialization by selection for wavelet network training. *Neurocomputing*, 34(1), 131-143.

32. Pakdin Amiri, A., Pakdin Amiri, M., & Pakdin Amiri, M. (2009). Introducing a Prediction Model in Total Stock Price Index Using Neural Networks Approach (Case Study: Tehran Stock Exchange). *Journal of Iran's Economic Essays*, 6(11), 83-108. (in Persian)

33. Pashootanizadeh, H., Ranaei Kordshouli, H., Abbasi, A., Moosavihaghighi, M. (2020). Simulation the Model of Effects of Behavioral and Macroeconomic Factors on the Tehran Stock Exchange Index with Using System Dynamics Approach. *Journal of Financial Management Perspective*, 10(29), 89-124.

34. Perez-Rodriguez, J. V., Torrab, S., & Andrada-Felixa, J. (2004). STAR and ANN models: Forecasting performance on the Spanish Ibex-35 stock index. *Journal of Empirical Finance*, 12(3), 490-509.

35. Qiu, M., & Song, Y. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. *PLoS ONE*, 11(5), 1-11.

36. Raee, R., Nikahd, A., & Habibi, M. (2017). The Index Prediction of Tehran Stock Exchange by Combining the Principal Components Analysis, Support Vector Regression and Particle Swarm Optimization. *Financial Management Strategy*, 4(15), 1-23. (in Persian)

37. Raoofi, A., & Mohammadi, T. (2018). Forecasting Tehran Stock Exchange Index Returns Using a Combination of Wavelet Decomposition and Adaptive Neural Fuzzy Inference Systems. *Iranian Journal of Economic Research*, 23(76), 107-136. (in Persian)

38. Salehi, M., & Garshasbi, F. (2019). Tehran Stock Exchange Index Forecasting Using Approach Adaptive Neural-Fuzzy Inference System and Imperialist Competitive Algorithm. *IT Management Studies*, 8(29), 5-34. (in Persian)

39. Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90, 1-63.

40. Singh, R., & Srivastava, S. (2017). Stock prediction using deep learning. *Multimedia Tools and Applications*, 76(18),18569-18584.

41. Sorzano, C. O. S., Vargas, J., & Pascual-Montano, A. (2014). A survey of dimensionality reduction techniques. *Cornell University Library Abstracts*, 1-35.

42. Tehrani, R., & Moradpour, S. (2012). Predicting the return of the stock market index using base radius artificial neural network models. *Financial Engineering and Securities Management*, 3(10), 75-90. (in Persian)

43. Tehrani, R., Heyrani, M., & Mansuri, S. (2019). A Comparison between Fama and French five-factor model and artificial neural networks in predicting the stock price. *Financial Engineering and Securities management(Portfolio Management)*, 10(39), 278-294. (in Persian)

44. Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*, 50(1), 258-269.

45. Ul Haq, A., Zeb, A., Lei, Z., & Zhang, D. (2021). Forecasting daily stock trend using multi-filter feature selection and deep learning. *Expert Systems With Applications*, 168(3),114444.

46. Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: A comparative. *Journal of Machine Learning Research*, 10(1-41), 66-71.

47.Wanga, D., & Zhao,Y.(2020). Using News to Predicton Investor Sentiment:Based on SVM Model, *Procedia Computer Science*, 174,191-199.

48. Yoo, P., Kim, M., & Jan, T. (2005). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In *Proceedings - International Conference on Computational Intelligence for Modelling, Control and Automation, (CIMCA 2005)*,(pp. 835-841). Piscataway, NJ: IEEE.

49. Zahedi, J., & Rounaghi, M. M. (2015). Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran Stock Exchange. *Physica A*, 438, 178-187.

50. Zhang, X. L., & Wu, J. (2013). Deep belief networks based voice activity detection. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(4), 697-710.

51. Zhong, X., & Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction, *Expert Systems With Applications*, 67,126-139.

52. Zolfaghari, M., Sahabi, B., & Bakhtyaran, M. (2020). Designing a Model for Forecasting the Stock Exchange Total Index Returns (Emphasizing on Combined Deep Learning Network Models and GARCH Family Models). *Financial Engineering and Securities management(Portfolio Management)*, 11(42), 138-171. (in Persian)

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی