



پیش‌بینی قیمت سهام با ترکیب XGBoost و بهینه‌ساز گرگ خاکستری باینری

هادی اسمعیلی^۱

شهاب جهانگیری^۲✉

علی رضازاده^۳

تاریخ پذیرش: ۱۴۰۴/۰۳/۱۰

تاریخ دریافت: ۱۴۰۴/۰۱/۰۵

چکیده

سرمایه‌گذاری کمی که توسط یادگیری ماشینی ایجاد می‌شود، فرصت‌های جدیدی را برای تولید بینش بیشتر از داده‌های مالی باز کرده است که منجر به توسعه ایده‌ها برای افزایش عملکرد سرمایه‌گذاری‌ها در بازارهای سهام می‌شود.

در این تحقیق، مدل معرفی شده به نام CBGWO-XGBoost برای پیش‌بینی قیمت پایانی روز آتی سهام تحت مطالعه پیشنهاد شده است. مدل پیشنهادی CBGWO را به عنوان بخش مهندسی ویژگی و XGBoost به عنوان پیش‌بینی کننده قیمت پایانی ترکیب می‌کند. این مدل با قیمت سهام شرکت‌های بورس اوراق بهادار تهران در قالب دو مجموعه ویژگی‌های توسعه یافته آموزش داده شده و سپس از مدل برای پیش‌بینی قیمت پایانی آتی سهام استفاده شده است. پس از آموزش و پیش‌بینی، ارزش واقعی و پیش‌بینی شده سهام با هم مقایسه می‌شود. به طور کلی مجموعه ویژگی‌های توسعه یافته ۲ عملکرد بهتری از مجموعه ویژگی‌های توسعه یافته ۱ بدون فرآیند مهندسی ویژگی داشته است. همچنین با اعمال فرآیند مهندسی ویژگی در مجموعه ویژگی‌های توسعه یافته ۲ الگوریتم‌ها به ترتیب CBGWO، BGWO2 و BGWO1 بهترین عملکرد را در نمادهای مختلف کسب نمودند. این مطالعه نشان داد که عملکرد موفقیت‌آمیز پیش‌بینی قیمت سهام به یک روش پیش‌بینی انحصاری بستگی ندارد. در عوض، نشان داده شده است که یک عملکرد پیش‌بینی خوب تا حد زیادی به فرآیندهای مهندسی ویژگی بستگی دارد، که عمده‌ا با یک الگوریتم یادگیری ترکیب می‌شود، که می‌تواند تعادل و هماهنگی خوبی بین افزایش ابعاد و کاهش ابعاد ایجاد کند.

واژه‌های کلیدی: پیش‌بینی قیمت سهام، XGBoost، بهینه‌ساز گرگ خاکستری باینری

طبقه بندی: JEL: C61, C63, G13, G17

۱ گروه علوم اقتصادی، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. esmaeili_hadi93@yahoo.com

۲ گروه علوم اقتصادی، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. (نویسنده مسئول) kh.jahangiri@urmia.ac.ir

۳ گروه علوم اقتصادی، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. a.rezazadeh@urmia.ac.ir

۱- مقدمه

عملکرد بازار مالی یک کشور عامل تعیین کننده‌ای برای وضعیت کلی اقتصادی آن کشور است که به اقتصاددانان و کارشناسان مالی این امکان را می‌دهد تا سلامت اقتصادی فعلی کشور را بسنجند. در میان بازارهای مالی مختلف، بازار سهام به عنوان یک نیروی محرکه اصلی و برجسته است. در بازار سهام، نوسان‌های گسترده در همه حال موجب ورود و خروج سرمایه گردیده و اثرات این جابجایی بر اقتصاد کشورها می‌تواند به شدت پرمخاطره باشد (ابراهیمی، ۱۳۹۸: ۲۸۳). وضعیت اقتصادی یک کشور به طور مستقیم یا غیرمستقیم بر بخش‌هایی مانند مالی، کشاورزی، فلزات و بانکداری سرمایه‌گذاری و غیره تأثیر می‌گذارد. رشد این بخش‌ها به نوسانات آن‌ها بستگی دارد که از اصل اساسی عرضه و تقاضا پیروی می‌کند. تقاضا برای یک بخش خاص مستقیماً بر بازار سهام تأثیر می‌گذارد؛ با افزایش عرضه، معامله‌گران و مؤسسات مالی وادار به سرمایه‌گذاری در آن بخش یا سهام می‌شوند و قیمت‌ها افزایش می‌یابند. علاوه بر این، پرداخت منظم سود سهام به تولید سود و بازده سرمایه‌گذاری شده کمک می‌کند. همچنین شناسایی فرصت مناسب برای خرید و فروش سهام و دستیابی به بازده مورد نظر برای سرمایه‌گذاران ضروری است (اوتونگ و همکاران^۱، ۲۰۲۰: ۱-۲).

به دلیل ماهیت غیرقابل پیش‌بینی بازار سهام، دستیابی به بازده سرمایه‌گذاری برای افراد بسیار دشوار است. تحلیل‌های اولیه، بنیادی و تکنیکال رویکردهای رایجی برای درک روندهای بازار هستند (کومار و آنموژی^۲، ۲۰۱۴: ۲۸۴). اما به دلیل دخالت اندیکاتورهای تاخیری و عدم دقت پیش‌بینی، دارای محدودیت‌های ذاتی هستند. این انگیزه محققان را برای توسعه تکنیک‌های بهبودیافته برای پاسخ مناسب به سناریوهای مختلف بازار بر اساس مدل‌های یادگیری ماشین ایجاد کرده است. در عصر حاضر، یادگیری ماشین مزایای قابل توجهی را نسبت به تکنیک‌های سنتی مانند تحلیل تکنیکال و بنیادی ارائه می‌دهند. این الگوریتم‌ها با استفاده از قدرت یادگیری ماشینی و هوش مصنوعی، پیش‌بینی قیمت سهام و شاخص‌ها را تسهیل می‌کنند. یادگیری ماشینی به عنوان یک رویکرد اضافی در کنار تجزیه و تحلیل تکنیکال و بنیادی عمل می‌کند، با ترکیب این ابزارها یک پلت فرم معاملاتی قدرتمند را تشکیل می‌دهد. مدل‌های یادگیری ماشینی می‌توانند راه‌حلی برای مسائلی مانند پیش‌بینی و طبقه‌بندی قیمت سهام، مدیریت پرتفوی، معاملات الگوریتمی، تحلیل احساسات بازار سهام، ارزیابی ریسک و غیره ارائه دهند (سونکاوده و همکاران^۳، ۲۰۲۳: ۲).

علاقه افراد به بازار سهام در چند دهه اخیر رشد تصاعدی را تجربه کرده است. اگر یک شرکت کننده در بازار مانند یک سرمایه‌گذار خصوصی یا نهادی بتواند رفتار بازار را به طور دقیق پیش‌بینی کند، این امکان فراهم می‌شود که به طور مداوم بازدهی با ریسک بالاتری نسبت به بازار کسب کنند. این انگیزه استفاده از یادگیری ماشین و روش‌های هوش محاسباتی برای ایجاد مدل‌های دقیق برای پیش‌بینی بازار سهام است. در واقع، تعداد زیادی از مطالعات منتشر شده سعی در پیش‌بینی دقیق بازارهای سهام با توسعه مدل‌ها/سیستم‌های پیش‌بینی پیچیده

¹ Obthong et al.² Kumar & Thenmozhi.³ Sonkavde et al.

داشته‌اند و برخی از مطالعات گزارش داده‌اند که مدل‌های آن‌ها می‌تواند سود ایجاد کند. به طور کلی، پیش‌بینی بازار سهام به عنوان یکی از مرتبط‌ترین اما چالش برانگیزترین وظایف در تحقیقات مالی شناخته می‌شود. با این حال، توانایی یک سرمایه‌گذار برای دستیابی مداوم به بازده تعدیل شده با ریسک بالاتر از بازار می‌تواند در تضاد با فرضیه بازار کارا (EMH^1) باشد (کامبور و همکاران^۲، ۲۰۲۲: ۱).

با توجه به ویژگی‌های غیرخطی و غیر ثابت ذاتی داده‌های بازار سهام، پیش‌بینی قیمت‌ها یا روندهای آتی در بازارهای مختلف سهام دشوار است (سریجیرانون و همکاران^۳، ۲۰۲۲: ۲). طبق EMH ، "شکست دادن بازار" با پیش‌بینی بازده مالی آتی و کسب سود بیش از حد متوسط غیرممکن است. با این حال، در سال‌های اخیر، تکامل تکنیک‌های یادگیری ماشین، فرصت‌های جدیدی را برای کسب سود بیشتر با پیش‌بینی بهتر رفتار آینده سهام باز کرده است. اکثر مطالعات به دو دسته اصلی تقسیم می‌شوند، پیش‌بینی ارزش بعدی برای قیمت سهام، نرخ بازده یا شاخص بازار به عنوان یک روش رگرسیون یا پیش‌بینی چگونگی نوسان یک سهام یا بازار خاص بر اساس داده‌های تاریخی. اولی یک کار پیش‌بینی است که عمدتاً برای پیش‌بینی شاخص بازار و روند قیمت دارایی‌ها، از جمله سهام، اوراق قرضه، رمزارزها و غیره استفاده می‌شود. با این حال، ارزش پیش‌بینی شده یک سیگنال اطلاعاتی قابل توجه است که می‌تواند توسط یک استراتژی معاملاتی الگوریتمی یا طرحواره انتخاب سبب مورد استفاده قرار گیرد. برای هر کار پیش‌بینی خاص، باید دانست که چگونه انتخاب ویژگی، استخراج ویژگی و مدل‌های یادگیری برای دستیابی به حداکثر عملکرد پیش‌بینی ترکیب شوند (دژکام و منظوری^۴، ۲۰۲۳: ۱).

در این مقاله، یک رویکرد جدید برای پیش‌بینی قیمت سهام با استفاده از پیش‌بینی کننده XGBoost و اجزای آبی استخراج شده از داده‌های قیمت سهام با استفاده از انواع مختلف بهینه‌ساز گرگ خاکستری باینری پیشنهاد شد. نوآوری این تحقیق توسعه یک فرآیند مهندسی ویژگی برای انتخاب ویژگی‌های ورودی بهینه با استفاده از انواع مختلف بهینه‌ساز گرگ خاکستری باینری در پیش‌بینی قیمت سهام با استفاده از پیش‌بینی کننده XGBoost می‌باشد. رویکرد پیشنهادی در بخش‌های بعدی بیشتر مورد بحث قرار می‌گیرد. بخش دوم پیشینه شاخص مرتبط سال‌های اخیر را مرور می‌کند. سپس، مطالب مرتبط با روش‌شناسی تحقیق در بخش سوم ارائه می‌شود. طرح کلی رویکرد پیشنهادی به طور کامل در این بخش تشریح می‌شود. خواننده می‌تواند نتایج را در بخش چهارم بیابد. مقاله با نتیجه‌گیری در بخش پنجم پایان می‌یابد.

۲- پیشینه پژوهش

در این بخش پیشینه پژوهش شاخص در زمینه پیش‌بینی قیمت سهام با استفاده از داده‌های بورس ایران و سایر کشورهای جهان بررسی می‌شود. نادم و همکاران (۱۴۰۳) در تحقیقی به بررسی مقایسه‌ای مدل یادگیری عمیق با طبقه‌بندی دوتایی و چندتایی جهت پیش‌بینی روند بازار سهام از طریق تشخیص الگوهای فراکتال مبتنی بر

¹ Efficient-market hypothesis

² Kumbure et al.

³ Srijiranon et al.

⁴ Dezhkam & Manzuri

تئوری امواج الیوت پرداختند. نتایج حاصل نشان دهنده‌ی صحت ۱۸٪ در تشخیص الگوها در حالت طبقه‌بندی چندتایی و صحت ۶۱٪ در حالت طبقه‌بندی دوتایی است. لذا توان مدل یادگیری عمیق در تشخیص الگوهای فراکتال امواج الیوت و در نتیجه پیش‌بینی روند بازار، در حالت طبقه‌بندی دوتایی به طور قابل توجهی نسبت به حالت طبقه‌بندی چندتایی بالاتر است. چوداری و تاکار^۱ (۲۰۲۳)، یک سیستم شبکه عصبی با ضریب یکپارچه انتخاب ویژگی مبتنی بر تغییرات برای پیش‌بینی قیمت و روند سهام ارائه نمودند. مجموعه‌ای از ویژگی‌های انتخاب شده برای مدل‌هایی مانند LSTM، BPNN، GRU و CNN برای پیش‌بینی قیمت سهام و روند اعمال شد. ژائو و یانگ^۲ (۲۰۲۳) یک چارچوب یکپارچه مبتنی بر یادگیری عمیق برای پیش‌بینی حرکت قیمت سهام ارائه نمودند. در این مقاله، یک مدل ترکیبی جدید SA-DLSTM برای پیش‌بینی بازار سهام و شبیه‌سازی معاملات با ترکیب یک شبکه عصبی کانولوشنال تقویت‌شده مبتنی بر احساسات (ECNN)، مدل‌های رمزگذار خودکار حذف نویز (DAE) و مدل LSTM پیشنهاد شده است.

دژکام و همکاران (۲۰۲۳)، به پیش‌بینی بازار سهام برای تشکیل سبد کارآمد با ترکیب XGBoost و تبدیل هیلبرت-هوانگ (HHT) همت گماشتند. توچایی و معینی^۳ (۲۰۲۳)، به ارزیابی عملکرد طبقه‌بندی‌کننده‌های تجمیعی در پیش‌بینی بازده سهام با استفاده از ویژگی‌های مؤثر همت گماشتند. در میان مدل‌های تجمیعی تقویت‌کننده، XGBoost و AdaBoost به ترتیب بهترین و بدترین عملکرد پیش‌بینی را ارائه نمود. سائیتیا و یوکراتاناساک^۴ (۲۰۲۳)، به پیش‌بینی حرکت سهام با استفاده از یادگیری ماشینی بر اساس شاخص‌های فنی و جستجوهای Google Trend در تایلند پرداختند. برای نرخ پیش‌بینی موفقیت و بازده سالانه، Random Forest و XGBoost تقریباً مشابه بودند.

شرما و سخاوات^۵ (۲۰۲۲) به بهینه‌سازی سبد سهام و پیش‌بینی بازده با ادغام DBF^۶ اصلاح شده و RNN^۷ پرداختند. کومبوره و همکاران (۲۰۲۲)، به بررسی ادبیات تکنیک‌ها و داده‌های یادگیری ماشین برای پیش‌بینی بازار سهام پرداختند. پارک و همکاران^۸ (۲۰۲۲)، به پیش‌بینی بازار سهام با استفاده از رویکرد چند وظیفه‌ای پرداختند که LSTM و چارچوب جنگل تصادفی (RF) را ادغام می‌کند. البحلی و همکاران^۹ (۲۰۲۲) یک مدل جدید DenseNet با رمزگذار خودکار (AEI-DNET) برای پیش‌بینی‌های بازار سهام با استفاده از شاخص‌های فنی سهام ارائه نمودند. نایک و آلبوکرکی^{۱۰} (۲۰۲۲) یک مدل تجمیعی مبتنی بر جستجوی بهینه‌سازی ترکیبی برای بهینه‌سازی سبد سهام و پیش‌بینی بازده در سرمایه‌گذاری تجاری با استفاده از ANN و XGBoost ارائه نمودند.

¹ Chaudhari & Thakkar

² Zhao & Yang

³ Toocheai & Moeini

⁴ Saetia & Yokrattanasak

⁵ Sharma & Shekhawat

⁶ Deep Belief Network

⁷ Recurrent Neural Network

⁸ Park

⁹ Albahli et al

¹⁰ Naik & Albuquerque

عابدینی و همکاران (۱۴۰۲)، بهینه‌سازی سبد سهام صنایع مختلف در بورس اوراق بهادار تهران با استفاده از گارچ متعامد را مورد مطالعه قرار دادند. بر اساس مدل میانگین-واریانس کلاسیک پویا (بعنوان مناسب‌ترین مدل)، بیشترین وزن در سبد سهام در دوره مورد مطالعه بترتیب مربوط به صنعت دارویی (۰/۶۳۳۶) و صنعت شیمیایی (۰/۳۵۳۹) بوده است. روحانی و همکاران (۱۴۰۲)، به تحلیل آثار شوک‌های سیاست پولی بر قیمت سهام در اقتصاد ایران؛ کاربردی ترکیبی از روش میانگین‌گیری بیزین و رهیافت خودرگرسیون برداری تعمیم‌یافته با پارامترهای متغیر در طول زمان پرداختند. طبق نتایج تغییرات یک انحراف معیار در شاخص سیاست پولی در طی زمان بر قیمت سهام افزایشی و به صورت U شکل بوده است. تغییرات یک انحراف معیار در شاخص سیاست پولی در در ابتدا تا اواخر دوره تأثیر مثبت و قوی بر قیمت سهام داشته است. فلاح مقدم بابایی کفاکی (۱۴۰۲)، به بررسی تاثیر ارزش معاملات خرد سهام بر شاخص کل بورس ایران پرداختند. هدف این پژوهش ارایه نگرشی جدید با استفاده از مدل‌های ریاضی و ابزارهای مالی، به بررسی تغییرات داده‌های ارزش معاملات خرد سهام و تاثیر آن بر روی شاخص کل بورس ایران می‌باشد. نتایج حاصل از این تحقیق موید این مطلب هستند که با افزایش یا کاهش شاخص کل بورس ایران شاهد روندی مشابه در ارزش معاملات خرد سهام هستیم. همچنین با استفاده از آزمون نمای هرست نشان دادیم که ارزش معاملات خرد سهام رونددار است.

فریدی و همکاران (۱۴۰۱)، به بهینه‌سازی سبد سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران بر اساس از روش‌های ترکیبی یادگیری ماشین تجمیعی دو سطحی و الگوریتم‌های فراابتکاری چندهدفه مبتنی بر رویکرد زمان‌سنجی بازار پرداختند. حیدری و امیری (۱۴۰۱)، به بررسی قدرت مدل‌های مبتنی بر هوش مصنوعی در پیش‌بینی روند قیمت سهام بورس اوراق بهادار تهران پرداختند. یافته‌ها نشان داد که مدل‌های مبتنی بر یادگیری عمیق عملکرد بهتری نسبت به سایر مدل‌ها دارند. باقری مزرعه و همکاران (۱۴۰۱)، به توسعه یک رویکرد یادگیری تجمیعی جدید برای انتخاب سبد سهام با استفاده از ماشین بردار پشتیبان چند کلاسه و الگوریتم ژنتیک پرداختند.

چن و همکاران^۱ (۲۰۲۱)، به بهینه‌سازی سبد سهام میانگین واریانس با استفاده از پیش‌بینی قیمت سهام مبتنی بر مدل ترکیبی از XGBoost با یک الگوریتم کرم شب تاب بهبود یافته (IFA) پرداختند. تاکار و چودهری (۲۰۲۱-a)، تحقیق مروری جامعی در مورد بهینه‌سازی سبد سهام، قیمت سهام و پیش‌بینی روند با استفاده از بهینه‌سازی ازدحام ذرات ارائه نمودند. تاکار و چودهری (۲۰۲۱-b)، به بررسی جامع در مورد نیازها، چالش‌ها و جهت‌گیری‌های آینده شبکه‌های عصبی عمیق برای بازار سهام پرداختند. یون و همکاران^۲ (۲۰۲۱)، به پیش‌بینی جهت قیمت سهام با استفاده از الگوریتم ترکیبی GA-XGBoost با فرآیند مهندسی ویژگی سه مرحله‌ای پرداختند. در نتیجه، این مطالعه به طور تجربی ثابت می‌کند که یک عملکرد پیش‌بینی موفق تا حد زیادی به ترکیبی عمدی از فرآیندهای مهندسی ویژگی با یک مدل یادگیری پایه بستگی دارد.

¹ Chen et al

² Yun et al

نبی‌پور و همکاران^۱ (۲۰۲۱)، به استفاده از یادگیری ماشین برای پیش‌بینی بازار سهام پرداختند. نتایج نشان از برتری LSTM بود. تاکار و چودهری (۲۰۲۱-c)، به مرور پژوهش‌های مرتبط با ادغام در اطلاعات، ویژگی‌ها و مدل‌ها برای پیش‌بینی در بازار سهام پرداختند. حسینی و همکاران (۲۰۲۱) به انتخاب بهینه سید سهام با استفاده از الگوریتم ترکیبی هوش جمعی سالپ و سینوس کسینوس و شبکه‌های عصبی پرداختند. ما و همکاران^۲ (۲۰۲۰)، به بررسی مدل‌های بهینه‌سازی سبد سهام مبتنی بر پیش‌بینی با استفاده از شبکه‌های عصبی عمیق پرداختند. راماموهان و همکاران^۳ (۲۰۲۰)، چارچوبی برای استفاده از خروجی‌های پیش‌بینی روند سهام در انتخاب سهام با شبکه‌های LSTM و بهینه‌سازی سبد سهام ارائه نمودند. ژانگ و همکاران^۴ (۲۰۲۰)، به بهینه‌سازی ردیابی شاخص سهام با استفاده از رمزگذارهای خودکار پرداختند. واچانی و همکاران^۵ (۲۰۲۰)، به مرور پیشینه تجزیه و تحلیل بازار سهام مبتنی بر یادگیری ماشین پرداختند.

شو و همکاران^۶ (۲۰۲۰)، به پیش‌بینی بازار سهام با ساختار یادگیری عمیق پشته شده با Bi-LSTM پرداختند. جیانگ و همکاران^۷ (۲۰۲۰)، چارچوب تعادل مجدد سبد سهام یکپارچه یادگیری ماشین با تنظیم ریسک-گریزی ارائه نمودند. نتایج تجربی نشان می‌دهد که مدل XGBoost بهترین پیش‌بینی حرکت بازار را ارائه می‌کند. حسینی و همکاران (۲۰۲۰)، به ارائه مدلی برای انتخاب سبد بهینه سهام با استفاده از الگوریتم هوش جمعی سالپ و شبکه‌های عصبی پرسپترون چندلایه پرداختند. وو و همکاران^۸ (۲۰۱۹)، به استفاده از یادگیری عمیق برای تصمیم‌گیری و بهینه‌سازی سرمایه‌گذاری‌ها و سبد دارایی مسئولیت‌پذیر اجتماعی پرداختند. باباجانی و همکاران (۱۳۹۸) به پیش‌بینی قیمت سهام در بورس اوراق بهادار تهران با استفاده از شبکه عصبی بهینه‌سازی شده با الگوریتم کلونی زنبور عسل مصنوعی پرداختند. دولو و حیدری (۱۳۹۶) به پیش‌بینی شاخص سهام با استفاده از ترکیب شبکه عصبی مصنوعی و مدل‌های فراابتکاری جستجوی هارمونی و الگوریتم ژنتیک پرداختند. نتایج حاصله نشان می‌دهد دقت پیش‌بینی مدل‌های فراابتکاری ژنتیک و جستجوی هارمونی در دوره آزمون بالاتر از شبکه عصبی عادی است.

با توجه پیشینه مورد بررسی در این تحقیق یک مدل ترکیبی از الگوریتم‌های یادگیری تجمیعی و فراابتکاری برای پیش‌بینی قیمت سهام توسعه داده می‌شود که در آن سه نوع الگوریتم فراابتکاری BGWO برای بهینه‌سازی ویژگی‌های ورودی الگوریتم XGBoost استفاده می‌شود.

¹ Nabipour et al

² Ma et al

³ Ramamohan et al

⁴ Zhang et al

⁵ Vachhani et al

⁶ Xu et al

⁷ Jiang et al

⁸ VO et al

شناسی ۳- روش

هدف اولیه این مقاله ساختن یک سیستم پیش‌بینی قیمت سهام با فرآیند مهندسی ۳ مرحله‌ای پیشرفته برای دستیابی به درجه بالایی از صحت پیش‌بینی است. مهندسی ویژگی یک رویکرد شناخته شده برای بهبود عملکرد مدل‌های یادگیری ماشین است. این یک پیش‌پردازش ساخت ویژگی‌های ورودی از داده‌های خام بر اساس دانش دامنه^۱ است. مهندسی ویژگی شامل تولید ویژگی، تبدیل ویژگی، استخراج ویژگی، انتخاب ویژگی و ارزیابی ویژگی است و اغلب به نوع داده و وابسته به کاربرد است (یان و همکاران، ۲۰۲۱: ۴). مهندسی ویژگی برای پیش‌بینی قیمت سهام معمولاً با ایجاد ویژگی‌های اضافی شروع می‌شود، زیرا داده‌های قیمت سهام معمولاً فقط شامل قیمت‌های تاریخی (باز، پایین‌ترین، بالاترین، و پایانی) و حجم معاملات هستند که با ویژگی‌های اضافی ترکیب می‌شوند تا عملکرد مدل را به میزان قابل توجهی افزایش دهند. مبنای منطقی پدیده بسط مجموعه ویژگی‌های برای بهبود عملکرد ممکن است همان «مزیت ابعاد» است، که به شدت برخی از مسائل یادگیری ماشینی با ابعاد بزرگ را ساده می‌کند. بنابراین، مجموعه ویژگی‌های تهیه شده از طریق (۱) گسترش مجموعه ویژگی‌ها با تولید شاخص‌های فنی، (۲) آماده‌سازی داده‌ها شامل پاکسازی و نرمال‌سازی داده‌ها (۳) انتخاب مجموعه ویژگی‌های بهینه با استفاده از الگوریتم ترکیبی BGWO-XGBoost مهندسی می‌شود. بعد از انتخاب ویژگی‌های بهینه از مدل XGBoost مبتنی بر مجموعه ویژگی‌های منتخب جهت پیش‌بینی قیمت استفاده می‌شود.

۳-۱- شرح پیش‌بینی قیمت سهام

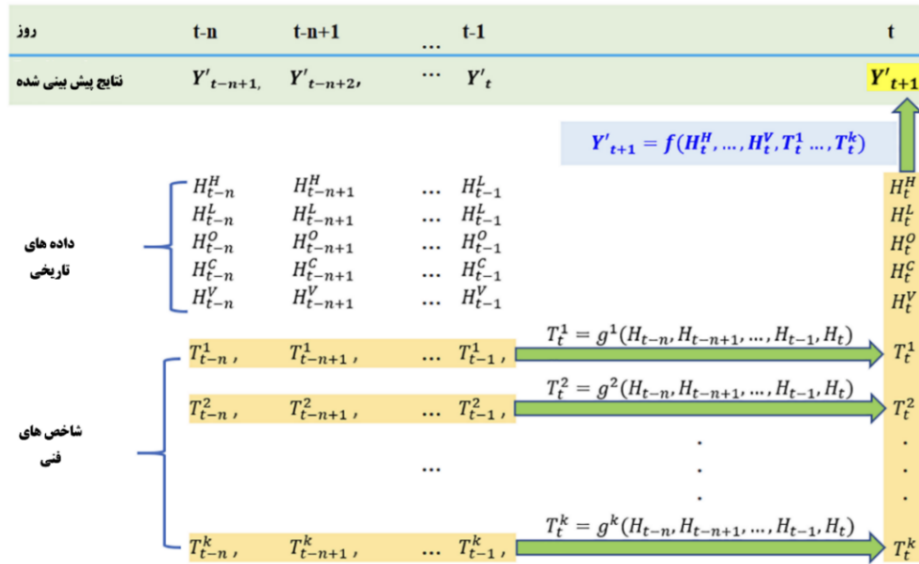
پیش‌بینی قیمت سهام شامل یک مدل پیش‌بینی است که در آن Y'_{t+1} قیمت پیش‌بینی شده پایانی روز $t + 1$ ، برای ورودی X_t ، مجموعه‌ای از ویژگی‌های ورودی روز t است و سپس با نتیجه واقعی Y_{t+1} یک مقدار عددی صحیح مقایسه می‌شود. نتیجه پیش‌بینی شده Y'_{t+1} تابعی از مجموعه ویژگی‌های ورودی X_t است. به این معنا که،

$$Y'_{t+1} = f(X_t) \quad (1)$$

که در آن f یک تابع غیرخطی است که مجموعه‌ای از ویژگی‌های ورودی X_t را در روز t به مقادیر نتیجه عدد صحیح در روز $t + 1$ همان‌طور که در شکل ۱ نشان داده شده است، ترسیم می‌کند.

¹ Domain Knowledge

² the blessing of dimensionality



شکل ۱- تصویری از فرآیند کلی پیش‌بینی قیمت سهام

منبع: یافته‌های پژوهشگر

سری زمانی قیمت سهام X_t یک فرآیند مارکوف از مرتبه n است که تحت تأثیر n حالت قبلی قیمت‌ها قرار می‌گیرد که می‌تواند به صورت زیر بیان شود:

$$p(x_t | x_{t-1}, x_{t-2}, \dots, x_1) = p(x_t | x_{t-1}, x_{t-2}, \dots, x_{t-n}) \text{ for } t > n \quad (2)$$

که در آن $p(A|B)$ یک احتمال شرطی از رویداد A را نشان می‌دهد. این سری زمانی قیمت سهام X_t از فرآیندهای کلاسیک مارکوف متفاوت است.

$$p(x_t | x_{t-1}, x_{t-2}, \dots, x_1) = p(x_t | x_{t-1}) \quad (3)$$

با توجه به اینکه حالت‌های گذشته، حال و آینده همه مستقل هستند. در نتیجه، حالت بعدی در فرآیند کلاسیک مارکوف فقط تحت تأثیر وضعیت فعلی است و نه از توالی حالت‌های قبلی. با این حال، در یک سری زمانی قیمت سهام با فرآیند مارکوف با حافظه n ، قیمت‌های روزانه سهام به قیمت‌های سهام در n حالت گذشته بستگی دارد. از منظر ویژگی مارکوفین داده‌های قیمت سهام، سری زمانی قیمت سهام شامل ۵ ویژگی ورودی: «بالا»، «کم»، «باز»، «پایانی» و «حجم» به یک روز خاص t برای پیش‌بینی قیمت سهام در روز بعد محدود می‌شود. از این رو، شاخص‌های فنی اضافی از این پنج مقدار قیمت تاریخی پایه تولید می‌شود تا از اطلاعات قیمت n حالت گذشته همان‌طور که در شکل ۱ نشان داده شده است استفاده شود.

مقادیر تاریخی اصلی: $H_t^V, H_t^C, H_t^O, H_t^L, H_t^H$ و H_t^V مقادیر قیمت روز جاری را نشان می‌دهند: به ترتیب «بالا»، «کم»، «باز»، «پایانی» و «حجم» و k شاخص فنی $T_t^1, T_t^2, \dots, T_t^k$ مقدار توابع غیرخطی مقادیر تاریخی n روز گذشته $H_t, H_{t-1}, \dots, H_{t-n}, H_{t-n+1}$ از روز t را نشان می‌دهد.

در نتیجه، نتیجه پیش‌بینی شده روز $t + 1$ را می‌توان به‌عنوان یک تابع غیرخطی از ویژگی‌های ورودی روز t که شامل مقادیر تاریخی $H_t^V, H_t^C, H_t^O, H_t^L, H_t^H$ و k شاخص‌های فنی $T_t^1, T_t^2, \dots, T_t^k$ که اطلاعات تاریخی روزهای $t-1, t, \dots, t-n, t-n+1$ را منعکس می‌کند و بنابراین معادله ۱ را می‌توان با معادله ۴ جایگزین کرد.

$$Y_{t+1} = f(H_t^H, H_t^L, H_t^O, H_t^C, H_t^V, H_t^1, H_t^2, \dots, T_t^K) \quad (۴)$$

که در آن عبارتند از «بالا»، «کم»، «باز»، «پایانی»، «حجم» و $T_t^1, T_t^2, \dots, T_t^k$ عبارتند از k شاخص فنی روز t هستند. در اینجا، شاخص فنی z ام روز t برابر با:

$$T_t^j = g^j(H_{t-n}, H_{t-n+1}, \dots, H_{t-1}, H_t) \quad (۵)$$

که در آن g^j یک تابع غیرخطی از مقادیر تاریخی n روز گذشته: $H_t, H_{t-1}, \dots, H_{t-n}, H_{t-n+1}$ از روز t است.

جدول ۱- شاخص‌های فنی رایج در مطالعات پیش‌بینی قیمت سهام

شاخص فنی	فرمول
مومنتوم	$C_t - C_{t-14}$
شاخص قدرت نسبی	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} Up_{t-i} / 14) / (\sum_{i=0}^{n-1} Dw_{t-i} / 14)}$
شاخص کانال کالا	$\frac{M_t - m_t}{d_t} \times 100$
اندیکاتور تصادفی K/.	$\frac{C_t - LL_{t-14}}{HH_{t-14} - LL_{t-14}} \times 100$
اندیکاتور لری ویلیام %R	$\frac{H_{14} - C_t}{H_{14} - L_{14}} \times 100$
میانگین متحرک واگرایی همگرایی	$EMA_{12} - EMA_{26}$
نوسانگر انباشت/توزیع	$\frac{H_t - C_{t-1}}{H_t - L_t} \times 100$

(یان و همکاران، ۲۰۲۱: ۵)

۳-۲- گسترش مجموعه ویژگی‌ها با تولید شاخص‌های فنی

اولین گام از مهندسی ویژگی ۳ مرحله‌ای پیشنهادی؛ گسترش مجموعه ویژگی‌ها با تولید شاخص‌های فنی است. دو دلیل برای استفاده از شاخص‌های فنی وجود دارد. اولین دلیل ادغام نشانگر قیمت سهام حالت‌های گذشته با قیمت وضعیت فعلی برای پیش‌بینی قیمت سهام در روز بعد بر اساس ویژگی حافظه n -حالت‌های مارکوبین همان‌طور که در ۳-۱ توضیح داده شده است. دوم استفاده از شاخص‌های فنی برای گسترش مجموعه ویژگی

ورودی برای گسترش ابعاد است. وقتی یک مجموعه داده ابعاد بالایی دارد، برخی از مسائل را می‌توان به سادگی به دلیل "مزیت ابعاد" حل کرد (کاینن^۱، ۱۹۹۷: ۲۹۱). این مطالعه با بهره‌گیری از ابعاد بالا، مجموعه ویژگی‌ها را با تولید شاخص‌های فنی گسترش می‌دهد. همچنین دامنه و تعداد شاخص‌های فنی که باید در پیش‌بینی قیمت سهام گنجانده شوند، باید مورد تحقیق بیشتر قرار گیرند، اما اکثر مطالعات قبلی که از شاخص‌های فنی برای پیش‌بینی قیمت سهام استفاده کرده‌اند، معمولاً از شاخص‌های فنی زیر استفاده می‌کردند که در جدول ۱ نشان داده شده است (چانگ و شین^۲، ۲۰۲۰: ۱۰؛ یان و همکاران، ۲۰۲۱: ۶). به همین دلیل، دو مجموعه ویژگی توسعه یافته ایجاد شد. مجموعه ویژگی‌های توسعه‌یافته ۱ از ۷ شاخص فنی رایج مطابق با جدول ۱ استفاده می‌کند و برای مقاصد مقایسه ایجاد شده است. توضیح مختصری از ۷ شاخص فنی رایج در زیر ارائه شده است (یان و همکاران، ۲۰۲۱: ۶)

- اندیکاتور مومنتوم: تفاوت بین قیمت‌های بسته شدن در روز t و روز $t - 14$ را اندازه‌گیری می‌کند تا افزایش (قدرت) یا کاهش (ضعف) قیمت سهام را تعیین کند.
- شاخص قدرت نسبی (RSI): تعیین می‌کند که آیا سهام بیش از حد خرید شده یا بیش از حد فروخته شده است. RSI از ۰ تا ۱۰۰ متغیر است. RSI نشان می‌دهد که سهام با نزدیک شدن به ۱۰۰ بیش از حد خرید شده است و همچنین نشان می‌دهد که سهام با نزدیک شدن به ۰ بیش از حد فروخته شده است.
- شاخص کانال کالا (CCI): تفاوت بین قیمت فعلی و میانگین تاریخی قیمت را اندازه‌گیری می‌کند. هنگامی که CCI بالای صفر است، نشان می‌دهد که قیمت بالاتر از میانگین تاریخی است و برعکس، زمانی که CCI زیر صفر است، قیمت کمتر از میانگین تاریخی است.
- اندیکاتور تصادفی %K: موقعیت قیمت بسته شدن یک سهم را در محدوده قیمت بالا و پایین در یک دوره معین به عنوان یک نوسانگر تصادفی اندازه‌گیری می‌کند. از ۰ تا ۱۰۰ متغیر است. اگر مقدار ۸۰ یا بالاتر باشد، محدوده خرید بیش از حد را نشان می‌دهد و اگر مقدار آن ۲۰ یا کمتر باشد، آنگاه فروش بیش از حد در نظر گرفته می‌شود.
- اندیکاتور لری ویلیام (LW) %R: سطح قیمت بسته شدن بازار را با بالاترین قیمت برای ۱۴ روز اندازه‌گیری می‌کند. از ۱۰۰- تا ۰- متغیر است. زمانی که ارزش آن بالاتر از ۲۰- باشد، سیگنال فروش و زمانی که ارزش آن کمتر از ۸۰- باشد، سیگنال خرید است.
- میانگین متحرک واگرایی همگرایی (MACD^۳): شاخصی که دو میانگین متحرک قیمت‌ها را با کم کردن میانگین متحرک نمایی (EMA^۴) ۲۶ روزه از میانگین متحرک نمایی ۱۲ روزه مقایسه می‌کند. این شاخص هنگامی که از بالای و پایین خط سیگنال خود عبور نماید به ترتیب سیگنال برای خرید و

¹ Kainen

² Chung & Shin

³ Moving Average Convergence Divergence

⁴ Exponential Moving Average

فروش می‌فرستد. سرعت عبور از خط سیگنال بیانگر این است که آیا بازار بیش از حد خرید شده است یا بیش از حد فروش شده است.

- نوسانگر انباشت/توزیع (A/D): یک شاخص تجمعی که از حجم و قیمت برای ارزیابی اینکه سهام در حال انباشت یا توزیع است، استفاده می‌کند. یک خط نوسانگر A/D افزایشی حاکی از روند صعودی قیمت است و یک خط A/D کاهشی حاکی از روند نزولی قیمت سهام است.

مجموعه ویژگی‌های توسعه‌یافته ۲ قصد دارد تا حداکثر تعداد ابعاد را استفاده کند و بنابراین ۸۷ اندیکاتور فنی از ۵ ویژگی ورودی اصلی تولید می‌شود که آن‌ها در جدول ۲ ارائه شده‌اند. بخشی اعظمی از اندیکاتورهای فنی (دسته دوم تا هفتم جدول ۲) در این مطالعه با استفاده از توابع ارائه‌شده توسط کتابخانه TA-Lib^۱ محاسبه می‌شوند. داده‌های بازار مالی هستند، استفاده می‌شود. همه اندیکاتورهای فنی تولید شده متعلق به یکی از ۷ گروه عملکردی شاخص هستند که با بررسی دانش تخصصی و مطالعات قبلی انتخاب می‌شوند (چن و همکاران، ۲۰۲۱: ۶؛ یان و همکاران، ۲۰۲۱: ۱۶-۱۷). آن‌ها مطالعات همپوشانی^۲، شاخص‌های مومنتوم، شاخص‌های حجم، شاخص‌های نوسان، تبدیل قیمت، شاخص‌های چرخه^۳ و بازده با تاخیر^۴ هستند که در جدول ۲ نشان داده شده است.

جدول ۲- شاخص‌های فنی مورد استفاده برای مجموعه ویژگی‌های توسعه‌یافته ۲

دسته‌بندی	شاخص‌های فنی	شرح
دسته اول: قیمت-های تاریخی	High	Highest Daily Price
	Low	Lowest Daily Price
	Open	Opening Price
	Close	Closing Price
	Volume	Trading Volume
دسته دوم: شاخص‌های مطالعات همپوشانی	1BBUPPER	Bollinger Bands Upper
	1BBMIDDLE	Bollinger Bands Middle
	1BBLOWER	Bollinger Bands Lower
	1DEMA	Double Exponential Moving Average
	1MIDPOINT	Midpoint over period
	1MIDPRICE	Midpoint Price over period
	1SMA	Simple Moving Average
	1T3	Triple Exponential Moving Average (T3)
	1TEMA	Triple Exponential Moving Average
	1TRIMA	Triangular Moving Average
	1WMA	Weighted Moving Average
	2ADX	Average Directional Movement Index
	2ADXR	Average Directional Movement Index Rating
دسته سوم: شاخص‌های مومنتوم	2APO	Absolute Price Oscillator
	2AROONDOWN	Aroon Down
	2AROONUP	Aroon Up

^۱ <https://github.com/TA-Lib/ta-lib-python>

^۲ Overlap Studies

^۳ Cycle Indicators

^۴ Lagged Return

دسته‌بندی	شاخص‌های فنی	شرح
	2AROONOSC	Aroon Oscillator
	2BOP	Balance of Power
	2CCI	Commodity Channel Index
	2CMO	Chande Momentum Oscillator
	2DX	Directional Movement Index
	2MACD	Moving Average Convergence/Divergence
	2MACDSIG	Moving Average Convergence/Divergence
	2MACDHIST	Moving Average Convergence/Divergence
	2MACDEXT	MACD with controllable MA type
	2MACDEXTSIG	MACD with controllable MA type
	2MACDEXTHIST	MACD with controllable MA type
	2MACDFIX	Moving Average Convergence/Divergence Fix 12/26
	2MACDFIXSIG	Moving Average Convergence/Divergence Fix 12/26
	2MACDFIXHIST	Moving Average Convergence/Divergence Fix 12/26
	2MFI	Money Flow Index
	2MINUS_DI	Minus Directional Indicator
	2MINUS_DM	Minus Directional Movement
	2MOM	Momentum
	2PLUS_DI	Plus Directional Indicator
	2PLUS_DM	Plus Directional Movement
	2PPO	Percentage Price Oscillator
	2ROC	Rate of change: ((price/prevPrice)-1)*100
	2ROCP	Rate of change Percentage: (price-prevPrice)/prevPrice
	2ROCR	Rate of change ratio: (price/prevPrice)
	2ROCR100	Rate of change ratio 100 scale: (price/prevPrice)*100
	2RSI	Relative Strength Index
	2SLOWK	Stochastic %K High
	2SLOWD	Stochastic %K Low
	2STOCHFH	Stochastic Fast High
	2STOCHFL	Stochastic Fast Low
	2STOK	Stochastic Relative Strength Index High
	2STOD	Stochastic Relative Strength Index Low
	2TRIX	1-day Rate-Of-Change (ROC) of a Triple Smooth EMA
	2ULTOSC	Ultimate Oscillator
	2WILLR	Williams' %R
دسته چهارم: شاخص‌های نوسان	3ATR	Average True Range
	3NATR	Normalized Average True Range
	3TRANGE	True Range
دسته پنجم: شاخص‌های حجم	4AD	Chaikin A/D Line
	4DOSC	Chaikin A/D Oscillator
دسته ششم: تبدیل قیمت	5AVGPRICE	Average Price
	5MEDPRICE	Median Price
	5TYPPRICE	Typical Price
	5WCLPRICE	Weighted Close Price
	6HT_DCPERIOD	Hilbert Transform - Dominant Cycle Period
	6HT_DCPHASE	Hilbert Transform - Dominant Cycle Phase
	6INPHASE	Hilbert Transform - Phasor Components Inpahse

دسته‌بندی	شاخص‌های فنی	شرح
دسته هفتم: شاخص‌های چرخه	6QUADRATURE	Hilbert Transform - Phasor Components Quadrature
	6SINE	Hilbert Transform - SineWave
	6LEADSINE	Hilbert Transform - LeadSineWave
	6HT_TRENDMODE	Hilbert Transform - Trend vs Cycle Mode
دسته هشتم: بازده با تاخیر	7LR_Ci_Ci_1	$\ln(\text{close price}_i / \text{close price}_{i-1})$
	7LR_Ci_1_Ci_2	$\ln(\text{close price}_{i-1} / \text{close price}_{i-2})$
	7LR_Ci_2_Ci_3	$\ln(\text{close price}_{i-2} / \text{close price}_{i-3})$
	7LR_Ci_3_Ci_4	$\ln(\text{close price}_{i-3} / \text{close price}_{i-4})$
	7LR_Hi_Oi	$\ln(\text{high price}_i / \text{open price}_i)$
	7LR_Hi_Oi_1	$\ln(\text{high price}_i / \text{open price}_{i-1})$
	7LR_Hi_Oi_2	$\ln(\text{high price}_i / \text{open price}_{i-2})$
	7LR_Hi_Oi_3	$\ln(\text{high price}_i / \text{open price}_{i-3})$
	7LR_Hi_1_Oi_1	$\ln(\text{high price}_{i-1} / \text{open price}_{i-1})$
	7LR_Hi_2_Oi_2	$\ln(\text{high price}_{i-2} / \text{open price}_{i-2})$
	7LR_Hi_3_Oi_3	$\ln(\text{high price}_{i-3} / \text{open price}_{i-3})$
	7LR_Li_Oi	$\ln(\text{low price}_i / \text{open price}_{i-3})$
	7LR_Li_1_Oi_1	$\ln(\text{low price}_{i-1} / \text{open price}_{i-1})$
	7LR_Li_2_Oi_2	$\ln(\text{low price}_{i-2} / \text{open price}_{i-2})$
	7LR_Li_3_Oi_3	$\ln(\text{low price}_{i-3} / \text{open price}_{i-3})$

(چن و همکاران، ۲۰۲۱: ۶؛ یان و همکاران، ۲۰۲۱: ۱۶-۱۷)

۳-۳- آماده‌سازی داده‌ها

همراه با گسترش فضای ویژگی با استفاده از شاخص‌های فنی، پیش‌پردازش داده‌ها برای مرحله پیش‌بینی اصلی زیر ضروری است. پیش‌پردازش داده‌ها از پاکسازی داده‌ها، تبدیل داده‌ها و نرمال‌سازی داده‌ها تشکیل شده است. برای مرحله پاکسازی داده‌ها، ویژگی‌های ناخواسته و غیرضروری از مجموعه ویژگی‌های توسعه یافته حذف می‌شوند. در این پژوهش مجموعه داده غیرضروری برای حذف وجود نداشت. تبدیل داده‌ها به عنوان یک مرحله پیش‌پردازش برای مرحله پیش‌بینی CBGWO-XGBoost مورد نیاز است. مدل‌های یادگیری ماشین، نمایش‌های مناسبی را برای داده‌های ورودی خود پیدا می‌کنند. تبدیل داده‌ها، آن‌ها را برای پردازش اصلی، مانند کار طبقه‌بندی، سازگارتر می‌کند (یان و همکاران، ۲۰۲۱: ۸). تمام مقادیر داده‌های ورودی به یک آرایه عددی همگن تبدیل می‌شوند تا به CBGWO-XGBoost وارد شوند. داده‌های تاریخی قیمت سهام دارای متغیرهای پیوسته با واحدهای اندازه‌گیری مختلف برای حجم و قیمت است. علاوه بر این، برخی از شاخص‌های فنی اضافه شده به داده‌های تاریخی اصلی، اندازه‌گیری نرخ هستند. نرمال‌سازی داده‌ها با استفاده از فرمول min-max در معادله ۶ برای پردازش ویژگی‌های مختلف مقیاس شده با هم مفید است. به طور خاص، نرمال‌سازی مقادیر ویژگی ورودی به گرادیان کاهشی کمک می‌کند تا خیلی سریع‌تر همگرا شود (چانگ و شین، ۲۰۱۸: ۱۰) و دقیقاً همه روابط را در داده‌ها حفظ می‌کند و در نتیجه هیچ گونه سوگیری ایجاد نمی‌کند (دوان و خاره^۱، ۲۰۲۰: ۵).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (۶)$$

¹ Devan & Khare

CBGWO-XGBoost-۳-۴- انتخاب ویژگی بهینه توسط الگوریتم

یک مجموعه داده با ابعاد بالا ممکن است باعث برآزش بیش از حد، افزایش غیرضروری مصرف حافظه، هزینه‌های محاسباتی گران و کاهش عملکرد یک الگوریتم یادگیری شود (لی و همکاران^۱، ۲۰۱۸: ۵-۳؛ هانگ و همکاران^۲، ۲۰۲۰: ۴۱۸۳). استخراج ویژگی^۳ و انتخاب ویژگی^۴ تکنیک‌های کاهش ابعاد محبوب برای غلبه بر مشکلات ناشی از داده‌ها با ابعاد بالا هستند. استخراج ویژگی، ویژگی‌های جدیدی را از ترکیب ویژگی‌های اصلی ایجاد می‌کند، اما قابلیت تفسیر و درک نسخه‌های اصلی را از دست می‌دهد. انتخاب ویژگی برخی از ویژگی‌های مهم را بدون از دست دادن قابلیت تفسیر حفظ می‌کند. فرآیند انتخاب ویژگی، ویژگی‌های نامربوط و زائد را حذف می‌کند، کارایی یادگیری را افزایش می‌دهد (یو و لیو^۵، ۲۰۰۳: ۱)، و دقت پیش‌بینی را بهبود می‌بخشد. روش‌های انتخاب ویژگی را می‌توان به روش‌های فیلتر، بسته‌بندی^۶ و تعبیه‌شده طبقه‌بندی کرد (یو و لیو، ۲۰۰۳: ۱؛ میائو و نیو^۷، ۲۰۱۶: ۹۲۰؛ رامیرز-گالگو و همکاران^۸، ۲۰۱۷: ۱۳۶؛ لی و همکاران، ۲۰۱۸: ۳-۵).

روش‌های فیلتر، ویژگی‌ها را بر اساس ویژگی‌های آماری هر ویژگی بدون استفاده از هیچ الگوریتم یادگیری انتخاب می‌کنند. از سوی دیگر، روش‌های بسته‌بندی و تعبیه‌شده، انتخاب ویژگی و الگوریتم یادگیری از پیش تعیین‌شده را برای انتخاب زیرمجموعه بهینه از ویژگی‌ها برای الگوریتم ترکیب می‌کنند. همچنین روش‌های فیلتر معمولاً محاسبات غیرتکراری را روی مجموعه داده اجرا می‌کنند و از نظر محاسباتی کارآمدتر از روش‌های بسته‌بندی یا تعبیه‌شده هستند. در همین حال، روش‌های بسته‌بندی یا تعبیه‌شده، تمامی ترکیب‌های ممکن از ویژگی‌ها را تشکیل می‌دهند و هر زیرمجموعه از ویژگی‌ها را برای بهترین عملکرد با الگوریتم یادگیری ارزیابی می‌کنند. از این رو، روش‌های بسته‌بندی یا تعبیه‌شده می‌توانند به انتخاب ویژگی با عملکرد بهتری دست یابند. روش بسته‌بندی با هدف اولیه این مطالعه که ارائه سیستمی برای پیش‌بینی بهتر قیمت سهام است، مطابقت دارد. الگوریتم ترکیبی پیشنهادی CBGWO-XGBoost یک روش انتخاب ویژگی مبتنی بر بسته‌بندی برای انتخاب یک مجموعه ویژگی بهینه برای طبقه‌بندی با حذف ویژگی‌های نامربوط و اضافی است.

به عنوان یک رویکرد ترکیبی یکپارچه، الگوریتم CBGWO-XGBoost با جمع‌آوری و ارتقاء مجموعه ویژگی‌ها برای دقت بهتر در طول نسل‌ها، مجموعه ویژگی‌های تقریباً بهینه یا بهینه را جستجو می‌کند. مجموعه ویژگی‌های بهینه دقت پیش‌بینی را با حداقل تعداد ویژگی به حداکثر می‌رساند. جمعیت اولیه دارای ۶۰ عضو است و هر عضو دارای همان تعداد ویژگی‌های باینری است که تعداد ویژگی‌های مربوطه در شکل ۲ توضیح داده شده است. پس از اینکه یک نسل توسط الگوریتم CBGWO تشکیل شد، سپس هر عضو در جمعیت توسط تابع بهینگی از پیش

¹ Li et al

² Huang et al

³ Feature Extraction

⁴ Feature Selection

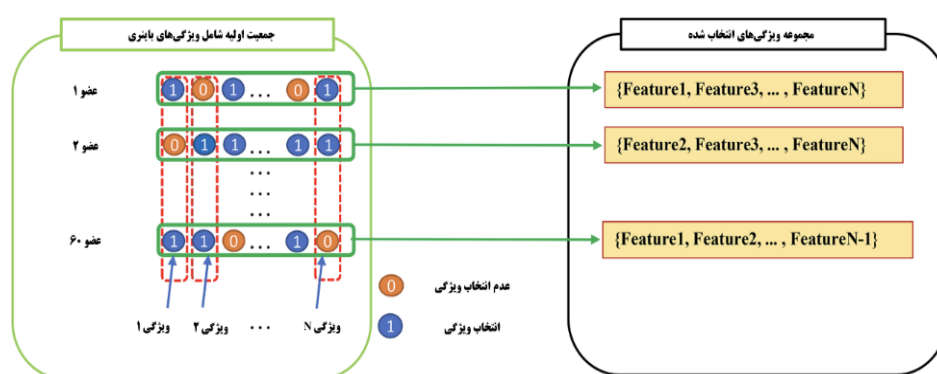
⁵ Yu & Liu

⁶ Wrapper

⁷ Miao & Niu

⁸ Ramírez-Gallego et al

تعریف شده طبقه‌بندی کننده XGBoost ارزیابی می‌شود. تابع بهینگی همان میزان میانگین مربعات خطا الگوریتم طبقه‌بندی کننده XGBoost می‌باشد. طبقه‌بندی کننده XGBoost مجموعه داده‌های آموزشی را با ویژگی‌های انتخاب شده آموزش می‌دهد، دقت طبقه‌بندی را با استفاده از هر مجموعه ویژگی اندازه‌گیری می‌کند، و دقت را با مجموعه ویژگی به فرآیند CBGWO برمی‌گرداند. در ادامه توضیحات مربوط به الگوریتم‌های CBGWO و XGBoost ارائه می‌شود.



شکل ۲- ساختار ویژگی‌های باینری در اعضای CBGWO-XGBoost

منبع: یافته‌های پژوهشگر

۳-۴-۱- بهینه‌ساز گرگ خاکستری

بهینه‌ساز گرگ خاکستری (GWO) یک روش بهینه‌سازی فراابتکاری است که توسط میرجلیلی و همکارانش در سال ۲۰۱۴ توسعه یافته است. به طور معمول، گرگ‌های خاکستری در یک گله با اندازه گروه ۵ تا ۱۲ زندگی می‌کنند. GWO مشخصه‌های صید و جستجو شکار گرگ‌های خاکستری در طبیعت را تقلید می‌کند. در GWO، جمعیت به آلفا، بتا، دلتا و امگا تقسیم می‌شوند. گرگ آلفا رهبر اصلی است که مسئول تصمیم‌گیری است. بتا گرگ دومین رهبر است که به آلفا در تصمیم‌گیری یا سایر فعالیت‌ها کمک می‌کند. گرگ دلتا به عنوان سومین رهبر گروه تعریف می‌شود که بر گرگ‌های امگا مسلط است (میرجلیلی و همکاران^۱، ۲۰۱۴: ۴۸-۴۹).

از نظر ریاضی، سه راه‌حل برتر در GWO به ترتیب آلفا (α)، بتا (β)، دلتا (δ) نامیده می‌شوند. فرض می‌شود بقیه امگا (ω) هستند. در GWO، روند شکار توسط آلفا، بتا، دلتا هدایت می‌شود، در حالی که امگا از این سه رهبر پیروی می‌کند. به طور کلی، GWO برای حل مسائل بهینه‌سازی پیوسته طراحی شده است. برای مسائل بهینه‌سازی باینری، مانند انتخاب ویژگی، یک نسخه باینری از GWO مورد نیاز است. اخیراً، امری^۲ و همکاران (۲۰۱۶) دو بهینه‌سازی جدید گرگ خاکستری باینری (BGWO1 و BGWO2) را برای مقابله با مسائل انتخاب ویژگی پیشنهاد

¹ Mirjalili et al

² Emary

کرد. در این تحقیق از این دو روش برای فرآیند مهندسی ویژگی استفاده خواهد شد. جزئیات مربوط به این روش‌ها در تحقیق امری و همکاران (۲۰۱۶) ارائه شده است.

۳-۴-۱-۱- بهینه‌ساز رقابتی گرگ خاکستری باینری

به طور کلی، BGWO در مقایسه با سایر بهینه‌سازی‌های فراابتکاری دارای مزایای ساده، انعطاف‌پذیر و سازگار است. با این حال، BGWO همچنین دارای محدودیت در رخداد بهینه محلی است. BGWO بهترین راه حل‌ها (رهبران) را در به‌روزرسانی موقعیت اعمال می‌کند، به این معنی که همه گرگ‌ها سعی می‌کنند به سمت موقعیت‌های رهبران حرکت کنند. به این ترتیب، گرگ‌ها به آرامی یا تقریباً مشابه رهبران می‌شوند. همه گرگ‌ها کم‌کم در بهینه محلی به دام می‌افتند. این منجر به تنوع کم و همگرایی زودرس خواهد شد. بنابراین، بهینه‌ساز رقابتی باینری گرگ خاکستری (CBGWO) برای رسیدگی به محدودیت BGWO در انتخاب ویژگی پیشنهاد می‌شود (تو و عبدالله^۱، ۲۰۲۱).

ایده کلی CBGWO از مفهوم رقابت بین هر زوج گرگ در جمعیت به ذهن خطور می‌کند. در CBGWO، گرگ‌ها به صورت تصادفی، دوتایی، از بین جمعیت، برای رقابت انتخاب می‌شوند. برای توضیح این مفهوم، N گرگ در جمعیت به طور تصادفی به $N/2$ زوج تقسیم می‌شوند که N تعداد گرگ‌های جمعیت است. پس از آن رقابت بین دو گرگ در هر زوج صورت می‌گیرد. این نشان می‌دهد که هر گرگ فقط یک بار در مسابقه شرکت می‌کند. از مسابقه، گرگ‌هایی که بهینه‌گی بهتری دارند، برنده نامیده می‌شوند. برعکس، گرگ‌هایی که در رقابت شکست می‌خورند، بازنده شناخته می‌شوند. برندگان مستقیماً بدون انجام به‌روزرسانی موقعیت به نسل بعدی منتقل می‌شوند. از طرف دیگر، بازندگان می‌توانند با یادگیری از برندگان، موقعیت خود را به روز کنند. به عبارت دیگر، تنها موقعیت $N/2$ گرگ در جمعیت به روز می‌شود (تو و عبدالله، ۲۰۲۱: ۳).

در این تحقیق از این روش به عنوان الگوریتم اصلی برای فرآیند مهندسی ویژگی استفاده خواهد شد. جزئیات مربوط به این روش‌ها در تحقیق تو و عبدالله (۲۰۲۱) ارائه شده است. شبه کد CBGWO را می‌توان در پیوست ۱ (الگوریتم ۱) یافت.

۳-۴-۱-۲- XGBoost

الگوریتم XGBoost مدلی است که بر اساس ادغام درخت طبقه‌بندی و رگرسیون است. مدل درختی با بالا بردن گرادیان در الگوریتم XGBoost بهینه شده است. XGBoost اجرای فرآیند مرتبط با تقویت گرادیان تعمیم یافته است که شامل اصطلاحی در حوزه منظم‌سازی می‌باشد. این ابزار برای مبارزه با برازش بیش از حد و همچنین پشتیبانی از توابع ضرر قابل تمایز استفاده می‌شود. در ادامه شبه‌کد الگوریتم XGBoost در الگوریتم ۲ ارائه شده است (لو و همکاران^۲، ۲۰۲۱: ۱۱۷۵).

¹ Too & Abdullah

² Luo et al

Algorithm 2. EXtreme Gradient Boosting Algorithm

Input: instance set of current nodes

Input: feature dimension

$J(p) = 0$

$$G = \sum_{i \in I} g_i \quad H = \sum_{i \in I} h_i$$

For $k = 1$ to n do

$G_L = 0 \quad H_L = 0$

For j in sorted do

$$G_L = G_L + g_j \quad H_L = H_L + h_j$$

$$G_R = G + G_L \quad H_R = H + H_L$$

$$score = \max(score, J(p))$$

End

End

Output: Split with max score

۳-۵- ارزیابی دقت

برای بررسی صحت روش‌های پیش‌بینی سهام، از چهار شاخص میانگین درصد مطلق خطا (MAPE)، میانگین مربعات خطا (MSE)، میانگین خطای مطلق (MAE) و ریشه میانگین مربعات خطا (RMSE) در این پژوهش استفاده شده است. MSE و RMSE اغلب نشان دهنده پراکندگی نتایج هستند، در حالی که MAE و MAPE نشان دهنده انحراف نتایج هستند. این شاخص‌ها به صورت زیر تعریف می‌شوند (چن و همکاران، ۲۰۲۱: ۷):

$$\begin{aligned} MAPE &= \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i}, \\ MSE &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \\ MAE &= \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \\ RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \end{aligned} \quad (7)$$

که در آن $\hat{y}_i$ نشان دهنده قیمت پیش‌بینی شده، y_i نشان دهنده قیمت واقعی و n تعداد کل نمونه‌ها است.

۴- یافته‌های پژوهش

۴-۱- داده‌های مورد استفاده و پیش پردازش آنها

داده‌های مورد استفاده در این تحقیق شامل ۲۷ نماد بازار بورس و اوراق بهادار ارائه شده در جدول ۳ می‌باشد. مشخصات مربوط به تعداد داده‌های هر نماد در جدول ۳ ارائه شده است. این داده‌ها مربوط به تاریخ ۲۰۱۱/۰۱/۰۱ تا ۲۰۲۲/۰۱/۰۱ می‌باشند.

جدول ۳- مشخصات نمادهای بورسی استفاده شده و تعداد داده‌های مربوطه

دسته‌بندی شاخص	نام نماد	تعداد داده‌ها	تعداد داده‌های حذف شده	تعداد داده‌های نهایی
متنوع مالی (۹) نماد	بانک ملت (وبملت)	۲۲۴۵	۱	۲۲۴۴
	بانک صادرات ایران (وبصادر)	۲۰۶۹	۰	۲۰۶۹
	بانک تجارت (وتجارت)	۲۲۰۱	۰	۲۲۰۱
	بانک پارسیان (وپارس)	۲۰۲۸	۰	۲۰۲۸
	بانک اقتصاد نوین (ونوین)	۲۳۳۹	۰	۲۳۳۹
	بانک سینا (وسینا)	۲۲۷۹	۰	۲۲۷۹
	بانک کار آفرین (وکار)	۲۳۲۸	۰	۲۳۲۸
	شرکت سرمایه‌گذاری صنعت بیمه (وبیمه)	۲۳۸۱	۰	۲۳۸۱
	سرمایه‌گذاری سپه (وسپه)	۲۵۷۳	۰	۲۵۷۳
نفت (۴) نماد	سرمایه‌گذاری صنعت نفت (ونفت)	۲۴۹۷	۰	۲۴۹۷
	پالایش نفت اصفهان (شپنا)	۲۰۳۷	۱	۲۰۳۶
	نفت پارس (شنفت)	۲۲۳۱	۲	۲۲۲۹
	نفت بهران (شبهرن)	۲۳۹۲	۱	۲۳۹۱
کانی‌های غیرفلزی (۸) نماد	کارخانجات تولیدی شیشه رازی (کرازی)	۲۰۹۸	۰	۲۰۹۸
	پارس سرام (کسرام)	۲۰۲۵	۱	۲۰۲۴
	سرامیک‌های صنعتی اردکان (کسرا)	۲۱۰۱	۱	۲۱۰۰
	سایپا شیشه (کسایپا)	۱۹۶۶	۰	۱۹۶۶
	پشم شیشه ایران (کپشیر)	۲۳۱۱	۱	۲۳۱۰
	فرآورده نسوز پارس (کفپارس)	۱۵۰۴	۱	۱۵۰۳
	فرآورده‌های نسوز آذر (کاذر)	۲۱۸۲	۰	۲۱۸۲
	شیشه و گاز (کگاز)	۲۱۴۸	۱	۲۱۴۷
فلزات اساسی (۷) نماد	فولاد مبارکه اصفهان (فولاد)	۲۴۷۸	۰	۲۴۷۸
	ملی صنایع مس ایران (فملی)	۲۴۶۹	۰	۲۴۶۹
	فولاد خوزستان (فخوز)	۲۳۵۲	۲	۲۳۵۰
	فولاد امیرکبیر کاشان (فجر)	۱۹۱۱	۱	۱۹۱۰
	سرمایه‌گذاری توکا فولاد (وتوکا)	۲۴۱۰	۰	۲۴۱۰
	فولاد خراسان (فخاس)	۲۲۳۱	۰	۲۲۳۱
مجموع	۲۷ نماد	۵۹۷۸۶	۱۳	۵۹۷۷۳

منبع: یافته‌های پژوهشگر

با بررسی نمودارهای تغییرات قیمت و حجم معاملات ۲۷ نماد مورد بررسی در برخی نقاط دارای شکاف‌های زمانی هستیم. این شکاف‌های زمانی در برخی نقاط با تغییرات شدید قیمت مواجه بوده‌اند. بدین منظور در بخش پیش‌پردازش داده‌ها این نقاط حذف شد. بدین منظور نقاط با قدر مطلق اختلاف "قیمت پایانی تعدیل شده" امروز با روز بعد بالاتر از ۲۰ هزار ریال فیلتر شد. بدین منظور ۱۳ سطر از کل ۵۹۷۸۶ داده حذف شد. این داده‌ها به دلیل نادر بودن (حدود ۰/۰۲ درصد از کل) و کاهش دقت کلی مدل پیش‌بینی حذف شدند. جدول ۳ به تفکیک تعداد داده‌های نهایی و حذف شده مربوط به هر نماد بورسی را نمایش می‌دهد. جدول ۴ آمار خلاصه تغییرات قیمت‌های پایانی را برای ۲۷ نماد مورد بررسی در پژوهش حاضر را نشان می‌دهد.

جدول ۴- آمار خلاصه برای نمادهای بورسی استفاده شده

نام نماد	تعداد	میانگین	انحراف معیار	کمینه	بیشینه
بانک ملت (وبملت)	۲۲۴۴	۲۹۶۳/۴۱	۳۳۷۰/۵۳	۸۹۷	۲۷۱۶۰
بانک صادرات ایران (وبصادر)	۲۰۶۹	۱۳۲۰/۷۵	۷۷۸/۲۱	۴۰۲	۴۸۷۰
بانک تجارت (وتجارت)	۲۲۰۱	۱۵۲۵/۶۰	۸۶۰/۵۹	۳۹۲	۴۴۲۰
بانک پارسیان (ویپارس)	۲۰۲۸	۲۵۰۳/۴۳	۱۰۱۰/۱۱	۹۱۵	۶۳۲۱
بانک اقتصاد نوین (ونوین)	۲۳۳۹	۲۹۹۲/۴۴	۱۲۱۸/۴۲	۱۰۲۸	۹۶۶۸
بانک سینا (وسینا)	۲۲۷۹	۲۷۷۸/۸۲	۲۸۹۰/۹۰	۸۸۸	۲۱۵۹۰
بانک کار آفرین (وکار)	۲۳۲۸	۳۰۴۷/۰۳	۲۴۹۰/۶۴	۱۰۰۰	۱۹۱۱۰
شرکت سرمایه‌گذاری صنعت بیمه (وبیمه)	۲۳۸۱	۲۵۹۳/۷۸	۲۰۹۵/۹۴	۶۱۳	۱۲۲۹۰
سرمایه‌گذاری سپه (وسپه)	۲۵۷۳	۳۲۶۱/۴۴	۳۷۵۱/۹۲	۱۰۲۴	۲۲۹۰۰
سرمایه‌گذاری صنعت نفت (ونفت)	۲۴۹۷	۴۷۶۷/۶۵	۵۲۵۵/۶۳	۹۵۵	۲۷۰۵۰
پالایش نفت اصفهان (شینا)	۲۰۳۶	۹۲۹۴/۰۸	۸۸۰۱/۴۰	۱۹۱۱	۵۰۶۸۰
نفت پارس (شنفت)	۲۲۲۹	۱۷۰۵۴/۵۵	۱۸۹۳۳/۸۸	۳۰۳۴	۱۰۵۲۹۰
نفت تهران (شپهرن)	۲۳۹۱	۲۰۰۳۶/۸۸	۱۶۹۸۸/۴۶	۶۳۰۱	۱۰۵۳۶۰
کارخانجات تولیدی شیشه رازی (کراز)	۲۰۹۸	۶۷۶۸/۵۷	۷۰۶۲/۸۰	۱۱۳۸	۳۶۳۴۰
پارس سرام (کسرام)	۲۰۲۴	۱۵۷۹۱/۴۶	۳۳۰۳۰/۶۶	۱۴۴۱	۱۸۰۳۱۲
سرامیک‌های صنعتی اردکان (کسرا)	۲۱۰۰	۱۸۵۸۱/۴۶	۱۸۵۴۰/۸۰	۱۸۱۰	۸۳۳۷۳
سایپا شیشه (کسایپا)	۱۹۶۶	۷۶۵۷/۴۷	۱۰۶۸۵/۸۰	۷۸۸	۴۶۰۴۰
پشم شیشه ایران (کپشیر)	۲۳۱۰	۱۲۲۹۳/۹۵	۱۳۵۱۸/۸۴	۱۲۲۸	۶۷۴۹۰
فرآورده نسوز پارس (کفپارس)	۱۵۰۳	۱۵۴۸۸/۴۴	۳۶۵۱۷/۷۷	۱۲۳۰	۲۴۸۳۷۳
فرآورده‌های نسوز آذر (کاذر)	۲۱۸۲	۷۹۴۶/۹۴	۱۱۰۸۹/۴۳	۸۲۳	۵۳۳۷۰
شیشه و گاز (گگاز)	۲۱۴۷	۱۰۱۸۱/۲۵	۱۲۴۸۸/۳۰	۱۲۸۶	۶۹۸۳۰
فولاد مبارکه اصفهان (فولاد)	۲۴۷۸	۴۶۳۱/۳۰	۴۰۰۵/۷۱	۱۰۰۱	۲۳۶۴۰

نام نماد	تعداد	میانگین	انحراف معیار	کمینه	بیشینه
ملی صنایع مس ایران (فملی)	۲۴۶۹	۶۰۳۸/۴۰	۶۲۲۱/۵۱	۱۱۳۹	۳۷۴۱۰
فولاد خوزستان (فخوز)	۲۳۵۰	۱۲۶۰۷/۸۲	۱۱۰۷۷/۸۰	۱۸۶۷	۵۳۰۴۰
فولاد امیرکبیر کاشان (فجر)	۱۹۱۰	۱۶۱۷۵/۹۹	۱۷۹۷۱/۴۴	۱۰۰۲	۷۲۳۰۰
سرمایه‌گذاری توکا فولاد (توکا)	۲۴۱۰	۴۱۲۵/۱۳	۴۵۰۰/۹۰	۸۴۹	۳۲۹۶۰
فولاد خراسان (فخاس)	۲۲۳۱	۱۱۶۴۳/۳۴	۹۸۴۸/۱۳	۲۶۸۲	۵۳۵۵۰

منبع: یافته‌های پژوهشگر

بدون مهندسی ویژگی برای مجموعه داده اول XGBoost ۴-۲- نتایج الگوریتم

در این بخش نتایج مربوط به الگوریتم XGBoost بدون مهندسی ویژگی به تفکیک نمادهای بورسی ارائه می‌شود. بدین منظور برای ۲۷ نماد مورد بررسی به تفکیک پارامترهای بهینه و نتایج ارزیابی ارائه می‌شود. با توجه به اینکه آموزش مدل XGBoost دارای مجموعه‌ای از پارامترها می‌باشد. بدین منظور ابتدا مقادیر قابل اخذ این پارامترها در جدول ۵ ارائه شده است.

جدول ۵- مقادیر پارامترهای قابل تنظیم در الگوریتم XGBoost

پارامتر قابل تنظیم	مقادیر مربوطه
max_depth	(۸، ۱۰، ۱۲، ۱۵، ۱۸)
gamma	(۰/۰۰۱، ۰/۰۰۵، ۰/۰۱، ۰/۰۲)
learning_rate	(۰/۰۰۱، ۰/۰۰۵، ۰/۰۱، ۰/۰۵)
n_estimators	(۱۰۰، ۲۰۰، ۳۰۰، ۴۰۰، ۵۰۰)

منبع: یافته‌های پژوهشگر

داده‌های مورد استفاده در این تحقیق به صورت ۸۰٪ برای آموزش و ۲۰٪ برای آزمون تقسیم شد. با توجه به اینکه در آموزش مدل XGBoost از جستجوی جامع (با کمک GridSearchCV) بر روی مقادیر پارامتر مشخص شده (جدول ۵) برای مدل XGBoost برای تعیین مقادیر بهینه استفاده شد. GridSearchCV یک کتابخانه قدرتمند پایتون برای بهینه‌سازی هایپرپارامتر است. این کتابخانه کمک می‌کند تا از طریق فراپارامترهای از پیش تعریف شده مقادیر بهینه برای مدل مشخص شود. بنابراین، در پایان، می‌توان بهترین پارامترها را از فراپارامترهای لیست شده انتخاب کرد. جدول ۶ برای هر یک از ۲۷ نماد مورد بررسی مقادیر بهینه پارامترهای قابل تنظیم در الگوریتم XGBoost به همراه شاخص‌های ارزیابی دقت برای مجموعه داده اول (۱۲ ویژگی) ارائه می‌شود.

جدول ۶- مقادیر بهینه پارامترهای قابل تنظیم در الگوریتم XGBoost و نتایج شاخص‌های ارزیابی برای نمادهای

بورسی

نام نماد	max_depth	gamma	learning_rate	n_estimators	MSE	RMSE	MAE	MAPE	R ²
وبملت	۱۸	۰/۰۱	۰/۰۵	۱۰۰	۵۳۶۶۳/۷۱۸۱	۲۳۱/۶۵۴۳	۸۱/۸۵۳۴	۰/۰۲۰۱	۰/۹۹۶۵۴۲
وبصادر	۱۰	۰/۰۰۱	۰/۰۵	۱۰۰	۴۱۱۴/۶۳۰۳	۶۴/۱۴۵۴	۳۰/۹۴۹۷	۰/۰۲۲۹	۰/۹۹۳۵۴۹
وتجارت	۸	۰/۰۱	۰/۰۵	۴۰۰	۲۹۵۱/۵۴۳۸	۵۴/۳۲۸۱	۲۹/۵۲۰۱	۰/۰۱۷۲	۰/۹۹۵۵۶۰
وپارس	۸	۰/۰۱	۰/۰۵	۱۰۰	۱۰۱۵۲/۰۷۶۴	۱۰۰/۷۵۷۵	۵۸/۱۸۲۳	۰/۰۲۰۵	۰/۹۹۰۱۱۷
ونوین	۸	۰/۰۱	۰/۰۵	۲۰۰	۸۳۵۹/۷۸۸۵	۹۱/۴۳۱۹	۵۱/۳۶۰۷	۰/۰۱۴۴	۰/۹۹۳۴۵۱
وسینا	۸	۰/۰۱	۰/۰۵	۲۰۰	۲۸۶۹۰/۰۳۴۸	۱۶۹/۳۸۱۳	۶۸/۸۱۸۷	۰/۰۱۹۲	۰/۹۹۶۲۱۸
وکار	۸	۰/۰۱	۰/۰۵	۳۰۰	۴۷۲۹۰/۶۵۴۴	۲۱۷/۴۶۴۱	۶۱/۱۱۶۴	۰/۰۱۳۶	۰/۹۹۲۸۸۴
وبیمه	۱۸	۰/۰۱	۰/۰۵	۳۰۰	۳۸۵۰۳/۲۰۹۴	۱۹۶/۲۲۲۳	۶۹/۳۲۹۰	۰/۰۲۴۰	۰/۹۹۰۷۱۴
وسپه	۸	۰/۰۱	۰/۰۵	۳۰۰	۴۱۵۹۴/۷۰۲۴	۲۰۳/۹۴۷۸	۷۹/۸۷۱۱	۰/۰۱۷۵	۰/۹۹۷۱۸۱
ونفت	۸	۰/۰۱	۰/۰۵	۳۰۰	۹۰۶۲۳/۷۶۱۹	۳۰۱/۰۳۷۸	۱۳۸/۴۴۹۳	۰/۰۲۲۴	۰/۹۹۶۷۶۲
شپنا	۸	۰/۰۱	۰/۰۵	۴۰۰	۶۰۴۵۹۵/۸۸۳۳	۷۷۷/۵۵۷۶	۲۸۰/۲۹۵۲	۰/۰۲۷۶	۰/۹۹۱۴۷۶
شنفت	۸	۰/۰۱	۰/۰۵	۴۰۰	۵۱۶۸۳۴/۰۶۷۲	۷۱۸/۹۱۱۷	۳۰۴/۹۷۳۱	۰/۰۱۶۱	۰/۹۹۸۵۸۷
شبه‌رن	۸	۰/۰۱	۰/۰۵	۳۰۰	۹۸۸۵۸۹/۳۱۸۴	۹۹۴/۲۷۸۳	۴۲۳/۵۸۷۶	۰/۰۱۶۲	۰/۹۹۶۷۶۹
کرازای	۸	۰/۰۱	۰/۰۵	۳۰۰	۱۲۳۹۳۱/۲۷۷۵	۳۵۲/۰۳۸۷	۱۷۸/۶۷۵۶	۰/۰۲۴۴	۰/۹۹۷۶۰۱
کسرام	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۱۷۷۰۷۲۲/۴۹۷۶	۱۳۳۰/۶۸۵۰	۴۰۶/۵۸۰۴	۰/۰۲۲۹	۰/۹۹۸۰۴۰
کسرا	۸	۰/۰۱	۰/۰۵	۴۰۰	۱۰۰۴۱۱۸/۷۶۵۷	۱۰۰۲/۰۵۷۳	۵۵۷/۶۱۰۱	۰/۰۲۶۸	۰/۹۹۷۳۵۲
کسپا	۸	۰/۰۱	۰/۰۵	۳۰۰	۱۵۶۴۶۴/۸۲۴۹	۳۹۵/۵۵۶۳	۱۷۳/۶۸۸۷	۰/۰۲۴۸	۰/۹۹۸۴۵۰
کپشیر	۸	۰/۰۱	۰/۰۵	۳۰۰	۳۲۹۵۳۲/۷۹۲۷	۵۷۴/۰۴۹۵	۲۶۵/۷۱۱۲	۰/۰۲۴۵	۰/۹۹۸۱۴۸
کفپارس	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۴۲۱۶۲۲۳/۶۴۷۹	۲۰۵۳/۳۴۴۵	۶۳۱/۲۰۲۸	۰/۰۲۷۹	۰/۹۹۶۶۲۴
کاذر	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۸۱۳۹۵۲/۲۸۸۵	۹۰۲/۱۹۳۰	۲۴۶/۱۳۷۰	۰/۰۲۹۵	۰/۹۹۳۱۹۱
کگاز	۸	۰/۰۱	۰/۰۵	۳۰۰	۵۲۱۷۱۴/۲۲۱۵	۷۲۲/۲۹۷۹	۲۷۹/۴۲۱۵	۰/۰۲۳۶	۰/۹۹۶۰۸۲
فولاد	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۵۱۵۵۳/۷۱۴۹	۲۲۷/۰۵۴۴	۱۰۵/۵۷۰۰	۰/۰۱۷۹	۰/۹۹۷۰۸۱
فملی	۸	۰/۰۱	۰/۰۵	۴۰۰	۱۱۰۱۳۰/۲۳۲۲	۳۳۱/۸۵۸۸	۱۴۸/۵۰۹۵	۰/۰۱۸۱	۰/۹۹۷۷۱۹
فخوز	۸	۰/۰۱	۰/۰۵	۴۰۰	۳۲۹۵۱۷/۷۱۲۲	۵۷۴/۰۳۶۳	۲۶۵/۱۲۸۰	۰/۰۱۷۰	۰/۹۹۷۲۶۴
فجر	۸	۰/۰۱	۰/۰۵	۳۰۰	۷۸۷۷۱۶/۲۳۸۵	۸۸۷/۵۳۳۸	۴۱۳/۲۰۰۵	۰/۰۱۷۰	۰/۹۹۷۴۵۸
وتوکا	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۵۶۱۱۵/۰۵۶۴	۲۳۶/۸۸۶۲	۱۱۸/۳۳۳۸	۰/۰۲۴۷	۰/۹۹۷۱۳۲
فخاس	۸	۰/۰۱	۰/۰۵	۴۰۰	۳۷۵۹۴۹/۶۹۲۲	۶۱۳/۱۴۷۴	۱۳۴/۴۵۳۷	۰/۰۰۸۵	۰/۹۹۶۱۲۲

منبع: یافته‌های پژوهشگر

بدون مهندسی ویژگی برای مجموعه داده دوم XGBoost ۴-۳- نتایج الگوریتم

در این قسمت، نتایج مرتبط با الگوریتم XGBoost بدون استفاده از مهندسی ویژگی برای نمادهای بورسی به صورت جداگانه ارائه شده است. به این منظور، پارامترهای بهینه و نتایج ارزیابی برای ۲۷ نماد مورد بررسی به صورت جداگانه ارائه می‌شود. آموزش مدل XGBoost شامل مجموعه‌ای از پارامترها است. مقادیر این پارامترها در جدول ۵ آمده است. داده‌های استفاده شده در این تحقیق به نسبت ۸۰ به ۲۰ برای آموزش و آزمون تقسیم شده‌اند. جدول ۷ شامل مقادیر بهینه پارامترهای قابل تنظیم در الگوریتم XGBoost برای هر یک از ۲۷ نماد مورد بررسی به همراه شاخص‌های ارزیابی دقت برای مجموعه داده دوم (۸۷ ویژگی) است. مقایسه نتایج جداول ۶ و ۷ نشان می‌دهد که در کل، مجموعه داده دوم (۸۷ ویژگی) نسبت به مجموعه داده اول (۱۲ ویژگی) عملکرد بهتری داشته است. همچنین، نمادهای بورسی در دسته متنوع مالی نسبت به سایر دسته‌ها عملکرد بهتری داشته‌اند.

جدول ۷- مقادیر بهینه پارامترهای قابل تنظیم در الگوریتم XGBoost و نتایج شاخص‌های ارزیابی برای نمادهای

بورسی

نام نماد	max_depth	gamma	learning_rate	n_estimators	MSE	RMSE	MAE	MAPE	R ²
وبملت	۱۰	۰/۰۱	۰/۰۵	۴۰۰	۳۸۵۰۷/۴۹۲	۱۹۶/۲۳۳۳	۶۹/۷۲۵۴	۰/۰۱۷۶	۰/۹۹۷۵۱۹
وبصادر	۱۰	۰/۰۰۱	۰/۰۵	۱۰۰	۳۶۶۶/۱۱۵۴	۶۰/۵۴۸۵	۲۹/۳۹۰۳	۰/۰۲۲۵	۰/۹۹۴۲۵۲
وتجارت	۸	۰/۰۱	۰/۰۵	۳۰۰	۴۸۹۱/۶۲۵۷	۶۹/۹۴۰۲	۲۹/۲۷۷۱	۰/۰۱۷۶	۰/۹۹۲۶۴۱
ویپارس	۸	۰/۰۱	۰/۰۵	۳۰۰	۸۳۸۲/۳۴۹۷	۹۱/۵۵۵۲	۵۲/۲۳۰۲	۰/۰۱۸۹	۰/۹۹۱۸۴۰
ونوین	۱۰	۰/۰۱	۰/۰۵	۲۰۰	۹۱۱۶/۱۱۱	۹۵/۴۷۸۳	۴۹/۴۷۰۵	۰/۰۱۳۹	۰/۹۹۲۸۵۸
وسینا	۸	۰/۰۱	۰/۰۵	۱۰۰	۲۱۸۶۲/۴۶۶	۱۴۷/۸۵۹۶	۶۵/۵۶۴۸	۰/۰۱۹۹	۰/۹۹۷۱۱۸
وکار	۱۲	۰/۰۱	۰/۰۵	۳۰۰	۲۶۹۷۶/۹۶۲	۱۶۴/۲۴۶۶	۵۶/۷۲۷۹	۰/۰۱۳۷	۰/۹۹۵۹۴۱
وبیمه	۱۲	۰/۰۱	۰/۰۵	۳۰۰	۳۶۵۹۳/۶۰۵	۱۹۱/۲۹۴۵	۶۵/۵۴۷۸	۰/۰۲۲۵	۰/۹۹۱۱۷۵
وسپه	۸	۰/۰۱	۰/۰۵	۳۰۰	۵۵۲۸۵/۴۴۴	۲۳۵/۱۲۸۶	۸۹/۴۶۳۸	۰/۰۱۹۵	۰/۹۹۶۲۵۳
ونفت	۸	۰/۰۰۱	۰/۰۵	۳۰۰	۶۸۸۹۴/۹۹	۲۶۲/۴۷۸۶	۱۲۵/۸۹۵	۰/۰۲۱۵	۰/۹۹۷۵۳۹
شپنا	۸	۰/۰۰۱	۰/۰۵	۳۰۰	۲۲۸۹۸۳/۲	۴۷۸/۵۲۱۹	۲۴۰/۱۱۵۴	۰/۰۲۲	۰/۹۹۶۷۷۲
شغت	۸	۰/۰۱	۰/۰۵	۴۰۰	۶۴۲۹۳۵/۲۷	۸۰۱/۸۳۲۴	۳۲۰/۹۶۲۲	۰/۰۱۶۲	۰/۹۹۸۲۴۳
شپهرن	۸	۰/۰۱	۰/۰۵	۳۰۰	۷۰۵۰۲۳/۴۳	۸۳۹/۶۵۶۷	۳۸۱/۷۰۶۴	۰/۰۱۵۳	۰/۹۹۷۶۹۶
کرازای	۸	۰/۰۱	۰/۰۵	۳۰۰	۱۰۸۲۷۵/۰۴	۳۲۹/۰۵۱۷	۱۷۱/۸۰۷۳	۰/۰۲۳۲	۰/۹۹۷۹۰۴
کسرام	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۱۳۷۶۴۲۳/۷	۱۱۷۳/۲۱۰۹	۳۹۳/۶۳۳۴	۰/۰۲۲	۰/۹۹۸۴۷۷
کسرا	۸	۰/۰۰۱	۰/۰۵	۴۰۰	۹۳۴۵۴۰/۱۲	۹۶۶/۷۱۶۲	۵۳۷/۳۲۹۹	۰/۰۲۵۸	۰/۹۹۷۵۳۶
کسپا	۸	۰/۰۱	۰/۰۲	۳۰۰	۱۴۹۷۶۲/۹۸	۳۸۶/۹۹۲۲	۱۶۹/۱۶۹۹	۰/۰۲۴	۰/۹۹۸۵۱۶
کپشیر	۸	۰/۰۱	۰/۰۵	۳۰۰	۳۱۶۱۵۶/۸۳	۵۶۲/۲۷۸۲	۲۶۶/۴۷۲۴	۰/۰۲۴	۰/۹۹۸۲۲۳
کفپارس	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۳۵۵۹۲۰۳/۸	۱۸۸۶/۵۸۵۲	۵۹۸/۸۵۵	۰/۰۲۷۵	۰/۹۹۷۱۵۰

نام نماد	max_depth	gamma	learning_rate	n_estimators	MSE	RMSE	MAE	MAPE	R ²
کاذر	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۴۲۸۴۴۷/۸۶	۶۵۴/۵۵۹۳	۲۳۳/۵۷۳۴	۰/۰۲۷۴	۰/۹۹۶۴۱۶
کگاز	۸	۰/۰۱	۰/۰۵	۳۰۰	۴۸۸۵۴۴/۹۸	۶۹۸/۹۵۹۹	۲۶۶/۹	۰/۰۲۲۳	۰/۹۹۶۳۳۱
فولاد	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۴۳۶۱۰/۵۳۱	۲۰۸/۸۳۱۳	۹۹/۲۹۴	۰/۰۱۶۵	۰/۹۹۷۵۳۰
فملی	۸	۰/۰۱	۰/۰۵	۴۰۰	۱۲۲۹۵۹/۱	۳۵۰/۶۵۵۲	۱۵۱/۰۶۴۹	۰/۰۱۷۴	۰/۹۹۷۴۵۳
فخوز	۸	۰/۰۱	۰/۰۵	۴۰۰	۲۵۲۴۱۶/۹۸	۵۰۲/۴۱۱۲	۲۳۵/۳۷۹۱	۰/۰۱۵۸	۰/۹۹۷۹۰۴
فجر	۸	۰/۰۱	۰/۰۵	۳۰۰	۶۸۲۹۶۹/۷۹	۸۲۶/۴۱۹۹	۳۷۸/۷۱۷	۰/۰۱۶۱	۰/۹۹۷۷۹۶
وتوکا	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۷۰۸۶۱/۹۲	۲۶۶/۱۹۹	۱۱۸/۱۱۰۱	۰/۰۲۳۳	۰/۹۹۶۳۷۸
فخاس	۱۸	۰/۰۱	۰/۰۵	۴۰۰	۳۰۷۵۷۲/۷۵	۵۵۴/۵۹۲۴	۱۳۱/۵۸۸۱	۰/۰۰۹۱	۰/۹۹۶۸۲۸

منبع: یافته‌های پژوهشگر

برای مجموعه داده دوم CBGWO با مهندسی ویژگی توسط XGBoost-۴-۴ نتایج الگوریتم در این بخش نتایج مربوط به الگوریتم XGBoost با مهندسی ویژگی توسط CBGWO به تفکیک نمادهای بورسی ارائه می‌شود. مقادیر پارامترهای قابل تنظیم در CBGWO در جدول ۸ ارائه شده است. بدین منظور برای ۲۷ نماد مورد بررسی به تفکیک شاخص‌های بهینه انتخابی و نتایج ارزیابی ارائه می‌شود.

جدول ۸- مقادیر پارامترهای قابل تنظیم در الگوریتم CBGWO

پارامتر قابل تنظیم	مقادیر مربوطه
تعداد جمعیت اولیه	۶۰
ابعاد اعضای جمعیت	۸۷
تعداد تکرار	۱۰۰

منبع: یافته‌های پژوهشگر

نتایج شاخص‌های ارزیابی و ویژگی‌های بهینه انتخابی در الگوریتم CBGWO-XGBoost برای نمادهای بورسی به ترتیب در جدول ۹ و شکل ۳ ارائه شده است. نتایج اخذ شده نشان می‌دهد که به طور میانگین در نمادهای مختلف بورسی ۲۰/۶ ویژگی بهینه انتخاب شده است.

جدول ۹- نتایج شاخص‌های ارزیابی در الگوریتم CBGWO-XGBoost برای نمادهای بورسی

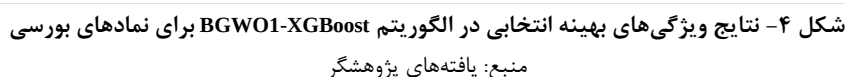
نام نماد	MSE	RMSE	MAE	MAPE	R ²
وبملت	۱۴۹۷۹/۷۳۰۸	۱۲۲/۳۹۱۷	۵۸/۳۵۴۴	۰/۰۱۸۰	۰/۹۹۹۰۳۵
وبصادر	۲۶۹۲/۶۹۸۷	۵۱/۸۹۱۲	۲۴/۹۰۳۱	۰/۰۲۰۷	۰/۹۹۵۷۷۸
وتجارت	۱۵۰۰/۸۵۱۶	۳۸/۷۴۰۸	۲۳/۵۷۴۳	۰/۰۱۵۰	۰/۹۹۷۷۴۲
ویپارس	۵۴۳۲/۲۴۶۹	۷۳/۷۰۳۸	۴۳/۶۵۹۷	۰/۰۱۶۷	۰/۹۹۴۷۱۲
ونوین	۵۷۹۸/۷۷۵۵	۷۶/۱۴۹۷	۴۲/۱۰۱۴	۰/۰۱۲۱	۰/۹۹۵۴۵۷
وسینا	۱۳۲۳۴/۱۲۰۱	۱۱۵/۰۳۹۶	۵۴/۲۷۰۰	۰/۰۱۸۸	۰/۹۹۸۲۵۶
وکار	۹۳۵۱/۱۴۷۷	۹۶/۷۰۱۳	۴۷/۲۷۹۳	۰/۰۱۳۲	۰/۹۹۸۵۹۳
وبیمه	۲۰۱۹۰/۹۴۰۷	۱۴۲/۰۹۴۸	۶۳/۲۴۶۸	۰/۰۲۳۰	۰/۹۹۵۱۳۱
وسپه	۱۸۷۳۶/۶۲۲۹	۱۳۶/۸۸۱۸	۶۴/۱۰۳۰	۰/۰۱۶۵	۰/۹۹۸۷۳۰
ونفت	۳۳۸۴۰/۸۳۷۰	۱۸۳/۹۵۸۸	۹۶/۳۶۱۵	۰/۰۱۸۸	۰/۹۹۸۷۹۱
شپنا	۸۵۵۹۰/۴۳۲۱	۲۹۲/۵۵۸۴	۱۷۵/۵۴۵۱	۰/۰۱۷۷	۰/۹۹۸۷۹۳
شنفت	۲۸۳۶۷۷/۱۶۹۰	۵۳۲/۶۱۳۵	۲۴۳/۳۳۷۹	۰/۰۱۴۵	۰/۹۹۹۲۲۵
شهرن	۳۷۹۹۳۰/۲۵۱۲	۶۱۶/۳۸۴۸	۳۱۴/۳۸۲۹	۰/۰۱۴۵	۰/۹۹۸۷۵۸
کرازلی	۶۲۲۹۷/۲۶۴۱	۲۴۹/۵۹۴۲	۱۴۱/۹۶۵۷	۰/۰۲۱۵	۰/۹۹۸۷۹۴
کسرام	۴۴۸۵۷۴/۷۵۵۴	۶۶۹/۷۵۷۲	۲۸۵/۲۳۸۵	۰/۰۲۱۹	۰/۹۹۹۵۰۴
کسرا	۷۱۶۴۰۲/۲۲۸۶	۸۴۶/۴۰۵۵	۴۷۹/۰۰۱۷	۰/۰۲۴۴	۰/۹۹۸۱۱۱
کساپا	۷۶۳۵۱/۱۱۸۹	۲۷۶/۳۱۷۱	۱۳۸/۹۴۴۳	۰/۰۲۳۱	۰/۹۹۹۲۴۳
کپشیر	۲۱۶۳۵۹/۸۰۷۴	۴۶۵/۱۴۴۹	۲۳۸/۲۳۱۹	۰/۰۲۳۵	۰/۹۹۸۷۸۴
کفپارس	۱۲۰۶۰۱۱/۲۸۵۰	۱۰۹۸/۱۸۵۵	۴۲۵/۰۹۳۴	۰/۰۲۸۴	۰/۹۹۹۰۳۴
کاذر	۱۴۰۷۶۰/۶۹۱۵	۳۷۵/۱۸۰۹	۱۶۹/۷۹۰۴	۰/۰۲۵۸	۰/۹۹۸۸۲۳
کگاز	۱۲۳۴۳۲/۱۲۹۰	۳۵۱/۳۲۹۱	۱۸۸/۱۲۳۸	۰/۰۲۰۸	۰/۹۹۹۰۷۳
فولاد	۲۴۳۸۸/۰۹۰۸	۱۵۶/۱۶۶۹	۸۵/۰۲۰۸	۰/۰۱۵۸	۰/۹۹۸۶۱۹
فملی	۶۰۰۶۱/۰۵۴۸	۲۴۵/۰۷۳۶	۱۲۵/۲۷۴۹	۰/۰۱۶۷	۰/۹۹۸۷۵۶
فخوز	۱۴۷۹۷۳/۲۱۱۹	۳۸۴/۶۷۲۹	۲۰۸/۶۰۸۸	۰/۰۱۵۰	۰/۹۹۸۷۷۱
فجر	۳۸۵۷۸۷/۹۶۵۲	۶۲۱/۱۱۸۳	۲۹۱/۶۲۷۶	۰/۰۱۳۸	۰/۹۹۸۷۵۵
وتوکا	۲۷۵۰۵/۵۹۶۹	۱۶۵/۸۴۸۱	۹۶/۲۳۹۷	۰/۰۲۲۵	۰/۹۹۸۵۹۴
فخاس	۱۰۲۴۷۲/۰۷۹۰	۳۲۰/۱۱۲۶	۱۲۸/۶۱۲۴	۰/۰۰۹۷	۰/۹۹۸۹۴۳

منبع: یافته‌های پژوهشگر

جدول ۱۰- نتایج شاخص‌های ارزیابی در الگوریتم BGWO1-XGBoost برای نمادهای بورسی

نام نماد	MSE	RMSE	MAE	MAPE	R ²
وبملت	۱۷۹۹۱/۲۹۵۹	۱۳۴/۱۳۱۶	۶۰/۰۸۵۴	۰/۰۱۷۵	۰/۹۹۸۸۴۱
وبصادر	۳۲۲۹/۰۹۲۰	۵۶/۸۲۵۱	۲۶/۵۶۵۴	۰/۰۲۰۷	۰/۹۹۴۹۳۷
وتجارت	۲۰۵۴/۲۹۷۷	۴۵/۳۲۴۴	۲۶/۴۰۹۵	۰/۰۱۶۲	۰/۹۹۶۹۱۰
ویپارس	۶۲۰۶/۸۸۵۷	۷۸/۷۸۳۸	۴۷/۲۷۸۹	۰/۰۱۷۵	۰/۹۹۳۹۵۷
ونوین	۶۹۵۹/۸۴۵۳	۸۳/۴۲۵۷	۴۶/۳۶۶۱	۰/۰۱۳۱	۰/۹۹۴۵۴۷
وسینا	۱۸۰۲۵/۸۲۷۳	۱۳۴/۲۶۰۳	۵۸/۲۷۸۸	۰/۰۱۷۹	۰/۹۹۷۶۲۴
وکار	۱۷۷۲۵/۲۴۵۰	۱۳۳/۱۳۶۲	۵۲/۷۸۸۶	۰/۰۱۳۵	۰/۹۹۷۳۳۳
وبیمه	۳۲۵۱۸/۳۵۰۷	۱۸۰/۳۲۸۵	۶۳/۲۲۰۳	۰/۰۲۲۶	۰/۹۹۲۱۵۸
وسپه	۳۲۶۷۴/۸۴۸۲	۱۸۰/۷۶۱۹	۷۱/۸۷۸۷	۰/۰۱۶۴	۰/۹۹۷۷۸۵
ونفت	۴۱۰۲۱/۵۲۹۵	۲۰۲/۵۳۷۷	۱۰۵/۸۹۹۳	۰/۰۱۹۶	۰/۹۹۸۵۳۴
شپنا	۱۱۰۷۵۲/۴۸۴۴	۳۳۲/۷۹۵۰	۱۹۵/۸۴۸۶	۰/۰۱۹۲	۰/۹۹۸۴۳۸
شنفت	۳۵۹۷۱۶/۱۱۳۱	۵۹۹/۷۶۳۴	۲۶۹/۴۱۵۴	۰/۰۱۴۹	۰/۹۹۹۰۱۷
شهرن	۴۹۰۱۲۵/۷۰۳۷	۷۰۰/۰۸۹۸	۳۳۲/۰۵۷۶	۰/۰۱۴۱	۰/۹۹۸۳۹۸
کرازی	۸۴۷۹۱/۷۲۰۱	۲۹۱/۱۹۰۲	۱۶۱/۴۰۱۵	۰/۰۲۲۸	۰/۹۹۸۳۵۸
کسرام	۷۵۱۲۱۱/۳۸۳۸	۸۶۶/۷۲۴۵	۳۳۴/۷۰۲۵	۰/۰۲۲۹	۰/۹۹۹۱۶۹
کسرا	۸۱۴۱۲۷/۲۳۳۰	۹۰۲/۲۹۰۰	۴۹۹/۳۲۹۹	۰/۰۲۵۱	۰/۹۹۷۸۵۳
کساپا	۱۰۹۸۰۵/۷۱۰۷	۳۳۱/۳۶۹۴	۱۴۵/۳۸۲۴	۰/۰۲۳۳	۰/۹۹۸۹۱۲
کپشیر	۲۵۸۲۴۱/۹۰۸۴	۵۰۸/۱۷۵۱	۲۴۷/۱۱۴۷	۰/۰۲۲۹	۰/۹۹۸۵۴۹
کفپارس	۲۱۳۳۹۰۸/۷۸۳۰	۱۴۶۰/۷۹۰۵	۴۸۳/۶۴۲۷	۰/۰۲۵۸	۰/۹۹۸۲۹۱
کاذر	۱۹۱۰۳۰/۵۴۲۰	۴۳۷/۰۷۰۴	۱۹۱/۵۱۷۵	۰/۰۲۶۵	۰/۹۹۸۴۰۲
کگاز	۱۷۷۲۳۵/۷۰۳۱	۴۲۰/۹۹۳۷	۲۱۳/۶۶۵۵	۰/۰۲۱۹	۰/۹۹۸۶۶۹
فولاد	۳۰۸۲۵/۶۱۷۸	۱۷۵/۵۷۲۳	۹۰/۸۷۳۳	۰/۰۱۶۱	۰/۹۹۸۲۵۴
فملی	۷۴۰۹۱/۵۷۲۳	۲۷۲/۱۹۷۷	۱۳۱/۸۹۱۴	۰/۰۱۶۷	۰/۹۹۸۴۶۵
فخوز	۲۰۴۲۴۶/۱۱۶۷	۴۵۱/۹۳۶۰	۲۲۲/۲۹۳۸	۰/۰۱۵۳	۰/۹۹۸۳۰۴
فجر	۴۹۵۵۵۱/۰۹۸۴	۷۰۳/۹۵۳۹	۳۴۱/۵۶۰۱	۰/۰۱۵۳	۰/۹۹۸۴۰۱
وتوکا	۳۸۸۲۵/۷۹۰۰	۱۹۷/۰۴۲۶	۱۰۴/۲۹۴۶	۰/۰۲۲۶	۰/۹۹۸۰۱۶
فخاس	۲۰۳۷۵۶/۲۲۸۱	۴۵۱/۳۹۳۷	۱۲۹/۱۱۰۷	۰/۰۰۹۲	۰/۹۹۷۸۹۸

منبع: یافته‌های پژوهشگر

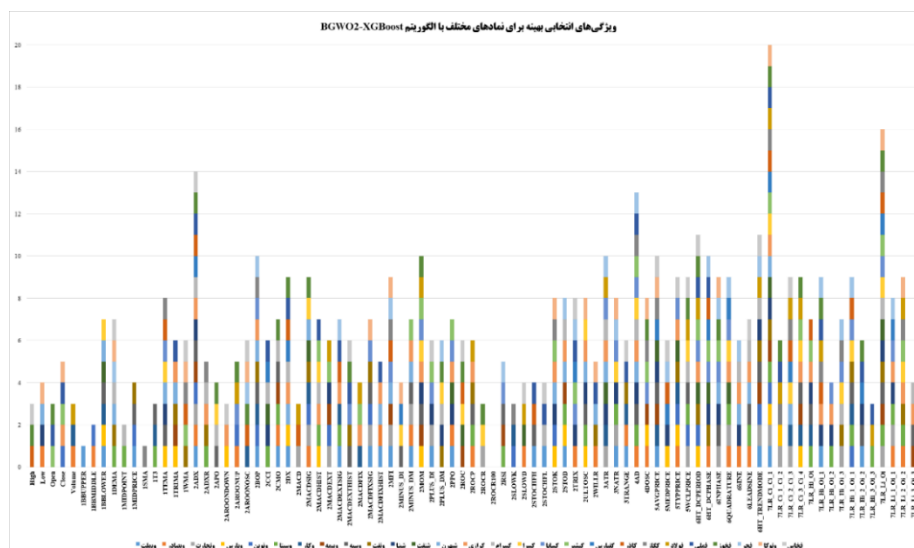


نتایج شاخص‌های ارزیابی و پارامترهای بهینه انتخاب شده در الگوریتم BGWO2-XGBoost برای نمادهای بورسی به ترتیب در جدول ۱۱ و شکل ۵ آمده است. نتایج به دست آمده نشان می‌دهد که به طور میانگین در نمادهای مختلف بورس، ۲۱/۴ ویژگی، بهینه انتخاب شده است.

جدول ۱۱- نتایج شاخص‌های ارزیابی در الگوریتم BGWO2-XGBoost برای نمادهای بورسی

نام نماد	MSE	RMSE	MAE	MAPE	R ²
وبملت	۱۵۰۹۸/۰۶۸۶	۱۲۲/۸۷۴۲	۵۸/۲۸۳۱	۰/۰۱۷۷	۰/۹۹۹۰۲۷
وبصادر	۲۸۹۳/۲۶۷۲	۵۳/۷۸۹۱	۲۵/۵۰۲۱	۰/۰۲۰۵	۰/۹۹۵۴۶۴
وتجارت	۱۶۵۴/۱۷۴۸	۴۰/۶۷۱۵	۲۴/۵۳۶۸	۰/۰۱۵۴	۰/۹۹۷۵۱۱
ویپارس	۵۷۲۸/۹۸۲۶	۷۵/۶۹۰۰	۴۵/۷۰۶۸	۰/۰۱۷۱	۰/۹۹۴۴۲۳
ونوین	۶۲۷۹/۷۷۵۹	۷۹/۲۴۵۰	۴۲/۸۳۷۱	۰/۰۱۲۳	۰/۹۹۵۰۸۰
وسینا	۱۴۱۴۳/۱۱۱۸	۱۱۸/۹۲۴۸	۵۴/۰۸۰۰	۰/۰۱۷۷	۰/۹۹۸۱۳۶
وکار	۹۶۵۹/۸۵۴۲	۹۸/۲۸۴۶	۴۶/۸۷۴۶	۰/۰۱۳۱	۰/۹۹۸۵۴۷
وبیمه	۲۵۲۹۳/۷۷۱۸	۱۵۹/۰۴۰۲	۶۶/۱۳۱۳	۰/۰۲۳۲	۰/۹۹۳۹۰۰
وسپه	۲۱۰۴۹/۱۲۰۸	۱۴۵/۰۸۳۲	۶۴/۴۴۳۹	۰/۰۱۶۰	۰/۹۹۸۵۷۳
ونفت	۳۴۰۳۱/۲۶۳۱	۱۸۴/۴۷۵۶	۹۸/۸۹۵۹	۰/۰۱۹۲	۰/۹۹۸۷۸۴
شپنا	۸۶۴۲۵/۳۵۵۰	۲۹۳/۹۸۱۹	۱۷۹/۱۲۰۲	۰/۰۱۸۰	۰/۹۹۸۷۸۱
شنفت	۲۹۲۱۲۷/۱۶۶۰	۵۴۰/۴۸۷۹	۲۵۷/۳۶۷۰	۰/۰۱۵۵	۰/۹۹۹۲۰۱
شهرن	۴۰۶۰۴۲/۹۶۵۸	۶۳۷/۲۱۵۰	۳۲۳/۷۸۲۷	۰/۰۱۴۷	۰/۹۹۸۶۷۳
کرازای	۷۱۲۹۰/۶۲۳۶	۲۶۷/۰۰۳۰	۱۴۷/۷۹۸۰	۰/۰۲۱۸	۰/۹۹۸۶۲۰
کسرام	۵۳۸۹۲۶/۰۰۷۱	۷۳۴/۱۱۵۸	۲۹۱/۱۵۲۲	۰/۰۲۱۸	۰/۹۹۹۴۰۴
کسرا	۷۴۵۵۳۷/۸۷۴۱	۸۶۳/۴۴۵۴	۴۸۳/۸۰۱۲	۰/۰۲۴۸	۰/۹۹۸۰۳۴
کساپا	۸۶۹۱۸/۱۷۵۰	۲۹۴/۸۱۸۹	۱۳۹/۶۳۰۳	۰/۰۲۲۸	۰/۹۹۹۱۳۹
کپشیر	۲۲۸۶۰۹/۳۱۸۱	۴۷۸/۱۳۱۱	۲۳۶/۸۵۷۳	۰/۰۲۳۰	۰/۹۹۸۷۱۵
کفپارس	۱۷۲۴۱۵۶/۹۲۸۱	۱۳۱۳/۰۷۱۶	۴۳۲/۲۷۸۳	۰/۰۲۵۵	۰/۹۹۸۶۲۰
کاذر	۱۵۳۶۴۷/۲۸۱۸	۳۹۱/۹۷۸۷	۱۷۶/۳۵۹۴	۰/۰۲۶۵	۰/۹۹۸۷۱۵
کگاز	۱۲۵۲۰۳/۰۹۳۲	۳۵۳/۸۴۰۵	۱۸۸/۳۹۷۳	۰/۰۲۱۰	۰/۹۹۹۰۶۰
فولاد	۲۵۹۵۳/۹۲۳۷	۱۶۱/۱۰۲۲	۸۶/۱۶۵۶	۰/۰۱۵۶	۰/۹۹۸۵۳۰
فملی	۶۳۳۸۶/۹۴۰۰	۲۵۱/۷۶۷۶	۱۲۴/۳۳۸۰	۰/۰۱۶۶	۰/۹۹۸۶۸۷
فخوز	۱۶۰۲۴۴/۲۸۶۲	۴۰۰/۳۰۵۲	۲۱۶/۸۲۳۰	۰/۰۱۵۵	۰/۹۹۸۶۶۹
فجر	۳۹۷۸۲۵/۸۳۶۱	۶۳۰/۷۳۴۴	۳۰۷/۹۹۴۹	۰/۰۱۵۰	۰/۹۹۸۷۱۶
وتوکا	۲۸۲۳۹/۱۶۸۵	۱۶۸/۰۴۵۱	۹۴/۳۳۳۹	۰/۰۲۲۲	۰/۹۹۸۵۵۷
فخاس	۱۵۷۴۲۰/۴۸۷۷	۳۹۶/۷۶۲۵	۱۱۸/۵۹۶۴	۰/۰۰۸۶	۰/۹۹۸۳۷۶

منبع: یافته‌های پژوهشگر



شکل ۵- نتایج ویژگی‌های بهینه انتخابی در الگوریتم BGWO2-XGBoost برای نمادهای بورسی
منبع: یافته‌های پژوهشگر

گیری ۵- نتیجه

وظایف یادگیری ماشین به شدت به کیفیت ویژگی‌های ورودی وابسته است. هنگام پیش‌بینی بازارهای سهام، عملکرد مدل نقش مهمی در سرمایه‌گذاری و استراتژی‌های معاملاتی سودآورتر بازی می‌کند. سری‌های زمانی مالی پر نویز هستند، در حالی که اطلاعات موجود در مجموعه برای ثبت روندهای آینده بسیار مهم است (دژاک و منظوری، ۲۰۲۳: ۱). تا آنجا که ما می‌دانیم، هیچ تحقیقی برای اعمال انواع مختلف بهینه‌سازی گرگ خاکستری باینری در پیش‌بینی قیمت باینری بعدی انجام نشده است.

این مطالعه یک سیستم پیش‌بینی قیمت سهام را با استفاده از الگوریتم BGWO-XGBoost با یک فرآیند مهندسی ویژگی ۳ مرحله‌ای پیشرفته پیشنهاد می‌کند. این مطالعه نشان می‌دهد که عملکرد موفقیت‌آمیز در پیش‌بینی قیمت سهام عمدتاً از فرآیند مهندسی ویژگی پیشنهادی است که از بسط مجموعه ویژگی‌ها، آماده‌سازی داده‌ها و انتخاب مجموعه ویژگی بهینه تشکیل شده است. به ویژه، مهم‌ترین افزایش دقت ناشی از گسترش ویژگی با افزودن ۸۲ شاخص فنی به داده‌های اولیه قیمت سهام تاریخی است. به طور کلی مجموعه ویژگی‌های توسعه یافته ۲ عملکرد بهتری از مجموعه ویژگی‌های توسعه یافته ۱ بدون فرآیند مهندسی ویژگی داشته است. همچنین با اعمال فرآیند مهندسی ویژگی در مجموعه ویژگی‌های توسعه یافته ۲ الگوریتم‌ها به ترتیب CBGWO2 و BGWO1 بهترین عملکرد را در نمادهای مختلف کسب نمودند. بنابراین، یافته‌های این تحقیق شواهد تجربی خوبی از پدید آمدن مزیت ابعاد (کاین، ۱۹۹۷) از طریق بسط مجموعه ویژگی‌ها ارائه می‌کند. یک دیگر از دستاوردهای

این مطالعه، تولید یک مجموعه ویژگی بهینه برای رهایی از نفرین ابعاد است. مجموعه ویژگی‌های بهینه داده‌های مربوط به نمادهای مختلف بورسی مورد بررسی در این مطالعه به طور میانگین تنها دارای ۲۱ ویژگی است که به طور قابل ملاحظه‌ای کمتر از ۸۷ ویژگی مجموعه ویژگی‌های توسعه یافته ۲ است، در حالی که دقت پیش‌بینی مورد نظر بهبود یافته است. به طور کلی، مجموعه‌ای از ویژگی‌های بهینه نسبت به مجموعه ویژگی‌های توسعه یافته ۲ مطلوب‌تر است، زیرا مدل با تعداد ویژگی‌های کمتر می‌تواند با هزینه محاسباتی کمتر، بیش‌برازش کمتر و قابلیت تفسیر آسان‌تر، عملکرد بهتری داشته باشد. به طور تجربی از طریق این مطالعه ثابت شد که عملکرد موفقیت آمیز پیش‌بینی قیمت سهام به یک روش پیش‌بینی انحصاری بستگی ندارد. در عوض، نشان داده شده است که یک عملکرد پیش‌بینی خوب تا حد زیادی به فرآیندهای مهندسی ویژگی بستگی دارد، که عمده‌ا با یک الگوریتم یادگیری پایه ترکیب می‌شود، که می‌تواند تعادل و هماهنگی خوبی بین افزایش ابعاد و کاهش ابعاد ایجاد کند.

فهرست منابع

- ابراهیمی، مهرزاد (۱۳۹۸). بررسی تاثیر متغیرهای کلان اقتصادی بر بازار سهام ایران با استفاده از الگوریتم‌های داده کاوی. فصلنامه اقتصاد مالی دانشکده اقتصاد دانشگاه تهران مرکزی، ۱۳(۴۹)، ۲۸۳-۳۰۹.
- باباجانی، جعفر؛ تقوا، محمدرضا؛ بولو، قاسم و عبدالهی، محسن (۱۳۹۸). پیش‌بینی قیمت سهام در بورس تهران با استفاده از شبکه عصبی بازگشتی بهینه شده با الگوریتم کلونی زنبور عسل مصنوعی. راهبرد مدیریت مالی، ۲۷(۲)، ۱۹۵-۲۲۸.
- باقری مزرعه، نسرين؛ دانشور، امیر و معدنچی زاج، مهدی. (۱۴۰۱). توسعه یک رویکرد جدید یادگیری جمعی برای انتخاب پورتفوی سهام با استفاده از ماشین بردار پشتیبان چند کلاسه و الگوریتم ژنتیک، مهندسی مالی و مدیریت اوراق بهادار، ۱۳(۵۰)، ۲۸۲-۳۰۵.
- حسینی، سیدعلی؛ پورزمانی، زهرا و جهانشاد، آریتا (۱۳۹۹). ارائه مدلی برای انتخاب سبد بهینه سهام با استفاده از الگوریتم هوش جمعی سالپ و شبکه‌های عصبی پرسپترون چندلایه، مهندسی مالی و مدیریت اوراق بهادار، ۱۱(۴۴)، ۴۷۹-۵۰۳.
- حیدری، مهدی؛ امیری، حمیدرضا (۱۴۰۱). بررسی قدرت مدل‌های مبتنی بر هوش مصنوعی در پیش‌بینی روند قیمت سهام بورس اوراق بهادار تهران، تحقیقات مالی، ۲۴، ۶۰۲-۶۲۳.
- دولو، مریم و حیدری، تکتیم (۲۰۱۷). پیش‌بینی شاخص سهام با استفاده از ترکیب شبکه عصبی مصنوعی و مدل‌های فرا ابتکاری جستجوی هارمونی و الگوریتم ژنتیک، فصلنامه اقتصاد مالی دانشکده اقتصاد دانشگاه تهران مرکزی، ۱۱(۴۰)، ۱-۲۴.
- روحانی، مریم؛ هوشمند، محمود؛ احمدی شامهری، محمدطاهر (۱۴۰۲). تحلیل آثار شوک‌های سیاست پولی بر قیمت سهام در اقتصاد ایران؛ کاربردی ترکیبی از روش میانگین‌گیری بیزین و رهیافت خودرگرسیون

- برداری تعمیم یافته با پارامترهای متغیر در طول زمان، فصلنامه اقتصاد مالی دانشکده اقتصاد دانشگاه تهران مرکزی، ۶۶(۱۸)، ۵۱-۶۸.
- عابدینی، سحر؛ ابونوری، اسمعیل و کشاورز حداد، غلامرضا (۱۴۰۲). بهینه‌سازی سید سهام صنایع مختلف در بورس اوراق بهادار تهران با استفاده از گارچ متعامد، فصلنامه اقتصاد مالی دانشکده اقتصاد دانشگاه تهران مرکزی، ۶۶(۱۸)، ۱-۲۶.
- فلاح مقدم، رضا و بابایی کفاکی، سامان (۱۴۰۲). بررسی تاثیر ارزش معاملات خرد سهام بر شاخص کل بورس ایران، فصلنامه اقتصاد مالی دانشکده اقتصاد دانشگاه تهران مرکزی، ۶۶(۱۸)، ۴۰۹-۴۳۴.
- نادم، مسعود؛ کامیابی، یحیی و ملکیان، اسفندیار (۱۴۰۳). بررسی مقایسه‌ای مدل یادگیری عمیق با طبقه‌بندی دوتایی و چندتایی جهت پیش‌بینی روند بازار سهام از طریق تشخیص الگوهای فراکتال مبتنی بر تئوری امواج الیوت، فصلنامه اقتصاد مالی دانشکده اقتصاد دانشگاه تهران مرکزی، ۶۶(۱۸)، ۱-۱۲۵.
- Albahli, S., Nazir, T., Mehmood, A., Irtaza, A., Alkhalifah, A., & Albattah, W. (2022). AEI-DNET: a novel densenet model with an autoencoder for the stock market predictions using stock technical indicators. *Electronics*, 11(4), 611.
- Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, 552-567.
- Chaudhari, K., & Thakkar, A. (2023). Neural network systems with an integrated coefficient of variation-based feature selection for stock price and trend prediction. *Expert Systems with Applications*, 219, 119527.
- Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean-variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing*, 100, 106943.
- Chung, H., & Shin, K. S. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. *Sustainability*, 10(10), 3765.
- Chung, H., & Shin, K. S. (2020). Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction. *Neural Computing and Applications*, 32, 7897-7914.
- Devan, P., & Khare, N. (2020). An efficient XGBoost-DNN-based classification model for network intrusion detection system. *Neural Computing and Applications*, 32, 12499-12514.
- Dezhkam, A., & Manzuri, M. T. (2023). Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert-Huangtransform. *Engineering Applications of Artificial Intelligence*, 118, 105626.
- Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, 172, 371-381.
- Huang, Q., Xia, T., Sun, H., Yamada, M., & Chang, Y. (2020). Unsupervised nonlinear feature selection from high-dimensional signed networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(4), 4182-4189.
- Jiang, Z., Ji, R., & Chang, K. C. (2020). A machine learning integrated portfolio rebalance framework with risk-aversion adjustment. *Journal of Risk and Financial Management*, 13(7), 155.
- Kainen, P. C. (1997). Utilizing geometric anomalies of high dimension: When complexity makes computation easier.
- Kumar, M., & Thenmozhi, M. (2014). Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. *International Journal of Banking, Accounting and Finance*, 5(3), 284-308.

- Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, 197, 116659.
- Li, B., Sun, Y., Aw, G., & Teo, K. L. (2019). Uncertain portfolio optimization problem under a minimax risk measure. *Applied Mathematical Modelling*, 76, 274-281.
- Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2018). Feature selection: A data perspective. *ACM computing surveys (CSUR)*, 50(6), 1-45.
- Luo, S., Zhang, S., & Cong, H. (2021, June). Research on consumer purchasing prediction based on xgboost algorithm. In *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)* (pp. 1173-1176). IEEE.
- Ma, Y., Han, R., & Wang, W. (2020). Prediction-based portfolio optimization models using deep neural networks. *IEEE Access*, 8, 115393-115405.
- Miao, J., & Niu, L. (2016). A survey on feature selection. *Procedia computer science*, 91, 919-926.
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in engineering software*, 69, 46-61.
- Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., & Salwana, E. (2020). Deep learning for stock market prediction. *Entropy*, 22(8), 840.
- Naik, M. J., & Albuquerque, A. L. (2022). Hybrid optimization search-based ensemble model for portfolio optimization and return prediction in business investment. *Progress in Artificial Intelligence*, 11(4), 315-331.
- Obthong, M., Tantisantiwong, N., Jeamwathanachai, W., & Wills, G. (2020). A survey on machine learning for stock price prediction: Algorithms and techniques.
- Park, H. J., Kim, Y., & Kim, H. Y. (2022). Stock market forecasting using a multi-task approach integrating long short-term memory and the random forest framework. *Applied Soft Computing*, 114, 108106.
- Qiu, M., & Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PloS one*, 11(5), e0155133.
- Ramamohan, V., Goyal, N., Parth, P., Mahlawat, S., & Prabhakar, U. (2020). A framework for utilizing stock trend prediction outputs in stock selection and portfolio optimization. Available at SSRN 3673874.
- Ramírez-Gallego, S., Lastra, I., Martínez-Rego, D., Bolón-Canedo, V., Benítez, J. M., Herrera, F., & Alonso-Betanzos, A. (2017). Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data. *International Journal of Intelligent Systems*, 32(2), 134-152.
- Saetia, K., & Yokrattanasak, J. (2023). Stock Movement Prediction Using Machine Learning Based on Technical Indicators and Google Trend Searches in Thailand. *International Journal of Financial Studies*, 11(1), 5.
- Sharma, M., & Shekhawat, H. S. (2022). Portfolio optimization and return prediction by integrating modified deep belief network and recurrent neural network. *Knowledge-Based Systems*, 250, 109024.
- Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D., & Bhat, S. K. (2023). Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. *International Journal of Financial Studies*, 11(3), 94.
- Srijiranon, K., Lertratanakham, Y., & Tanantong, T. (2022). A hybrid Framework Using PCA, EMD and LSTM methods for stock market price prediction with sentiment analysis. *Applied Sciences*, 12(21), 10823.

- Thakkar, A., & Chaudhari, K. (2021-a). A comprehensive survey on portfolio optimization, stock price and trend prediction using particle swarm optimization. *Archives of Computational Methods in Engineering*, 28(4), 2133-2164.
- Thakkar, A., & Chaudhari, K. (2021-b). A comprehensive survey on deep neural networks for stock market: the need, challenges, and future directions. *Expert Systems with Applications*, 177, 114800.
- Thakkar, A., & Chaudhari, K. (2021-c). Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions. *Information Fusion*, 65, 95-107.
- Too, J., & Abdullah, A. R. (2021). Opposition based competitive grey wolf optimizer for EMG feature selection. *Evolutionary Intelligence*, 14(4), 1691-1705.
- Toochaei, M. R., & Moeini, F. (2023). Evaluating the performance of ensemble classifiers in stock returns prediction using effective features. *Expert Systems with Applications*, 213, 119186.
- Vachhani, H., Obiadat, M. S., Thakkar, A., Shah, V., Sojitra, R., Bhatia, J., & Tanwar, S. (2020). Machine learning based stock market analysis: A short survey. In *Innovative Data Communication Technologies and Application: ICIDCA 2019* (pp. 12-26). Springer International Publishing.
- VO, N. N., He, X., Liu, S., & Xu, G. (2019). Deep learning for decision making and the optimization of socially responsible investments and portfolio. *Decision Support Systems*, 124, 113097.
- Wang, Q., Xu, W., & Zheng, H. (2018). Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. *Neurocomputing*, 299, 51-61.
- Xu, Y., Chhim, L., Zheng, B., & Nojima, Y. (2020). Stacked deep learning structure with bidirectional long-short term memory for stock market prediction. In *Neural Computing for Advanced Applications: First International Conference, NCAA 2020, Shenzhen, China, July 3–5, 2020, Proceedings 1* (pp. 447-460). Springer Singapore.
- Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 856-863).
- Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Systems with Applications*, 186, 115716.
- Zhang, C., Liang, S., Lyu, F., & Fang, L. (2020). Stock-index tracking optimization using auto-encoders. *Frontiers in Physics*, 8, 388.
- Zhao, Y., & Yang, G. (2023). Deep Learning-based Integrated Framework for stock price movement prediction. *Applied Soft Computing*, 133, 109921.

Stock Price Prediction with Combination of XGBoost and Binary Gray Wolf Optimizer

Hadi Ismaili¹
Shahab Jahangiri²
Ali Rezazadeh³

Received: 25/ March /2025 Accepted: 31/ May /2025

Abstract

Aim: Quantitative investing powered by machine learning has opened up new opportunities to generate more insights from financial data, leading to the development of ideas to enhance the performance of investments in stock markets.

Method: In this research, the proposed model called CBGWO-XGBoost is utilized to predict the closing price of the future stock under study. The proposed model combines CBGWO as the feature engineering part and XGBoost as the final price predictor. This model was trained with the stock prices of Tehran Stock Exchange companies in the form of two sets of developed features, and then the model was used to predict the future closing price of stocks. After training and forecasting, the actual and predicted stock values are compared.

Findings: In general, the developed feature set 2 performed better than the developed feature set 1 without the feature engineering process. Also, by applying the feature engineering process in the developed feature set 2, the algorithms CBGWO, BGWO2 and BGWO1 obtained the best performance, respectively.

Conclusion: This study showed that the successful performance of stock price forecasting does not depend on an exclusive forecasting method. Instead, it has been shown that a good prediction performance depends largely on feature engineering processes, deliberately combined with a training algorithm, which can strike a good balance between dimensionality increase and dimensionality reduction.

Keywords: Stock Price Prediction, XGBoost, Binary Gray Wolf Optimizer

JEL: C61, C63, G13, G17

¹ Department of Economic Sciences, Faculty of Economics and Management, Urmia University, Urmia, Iran.

² Department of Economic Sciences, Faculty of Economics and Management, Urmia University, Urmia, Iran.
(Corresponding author) kh.jahangiri@urmia.ac.ir

³ Department of Economic Sciences, Faculty of Economics and Management, Urmia University, Urmia, Iran.
a.rezazadeh@urmia.ac.ir