دوره هفدهم، شماره پاییز و زمستان ۱۴۰۳

مجله فناوری اطلاعات در طراحی مهندسی

Information Technology in Engineering Design

http://sanad.iau.ir/journal/ited

ارائه راهکار برای پیش‌بینی قیمت سهام به کمک تلفیق روش‌های تکامل تفاضلی و جنگل تصادفی

لادن ریاضی*

گروه مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، ایران *

(تاریخ دریافت: ۱۴۰۳/۰۵/۰۷ تاریخ پذیرش: ۱۴۰۳/۰۶/۲۵)

چکیده

پیش‌بینی قیمت سهام یک کار پیچیده است که برای قرن‌ها سرمایه‌گذاران و تحلیلگران مالی را مجذوب خود کرده است. با افزایش هوش مصنوعی و یادگیری ماشینی، محققان مدل‌های مختلفی را برای پیش‌بینی قیمت سهام، استفاده از داده‌های تاریخی و روند بازار توسعه داده‌اند. هدف این مدل‌ها شناسایی الگوها و همبستگی‌های بین شاخص‌های اقتصادی، اخبار بازار و قیمت سهام برای پیش‌بینی دقیق است. با وجود اینکه در زمینه‌ی پیش‌بینی قیمت سهام، تحقیقات گسترده‌ای صورت گرفته، ولی با این حال با افزایش استفاده از فناوری‌های نوین ضرروت انجام این فعالیت‌ها همچنان پابرجاست، در این مقاله روشی مبتنی بر تلفیق روش‌های تکامل تفاضلی و جنگل تصادفی برای پیش‌بینی قیمت سهام ارائه شده است. در روش پیشنهادی جنگل تصادفی می‌تواند به‌طور موثر قیمت‌های بازار سهام را با مدیریت و بررسی داده‌ها پیش‌بینی کند و روش تکامل تفاضلی با انتخاب بهترین مقادیر برای پارامترهای جنگل تصادفی به بهبود دقت پیش‌بینی‌های بازار سهام کمک می‌کند. نتایج حاصل از پیاده‌سازی این روش بر روی داده‌های شرکت AMD بیانگر این است که تا حدود میانگین ۷۴٪ نسبت به روش‌های جنگل تصادفی، شبکه عصبی و رگرسیون خطی دارای کاهش خطای آموزشی و همچنین تا حدود میانگین ۵۵٪ نسبت به روش‌های ذکرشده دارای کاهش خطای تست بوده است. این موضوع برتری و موثر بودن روش موردنظر نسبت به راهکارهای مقایسه‌شده را نشان می‌دهد.

کلمات کلیدی: جنگل تصادفی، تکامل تفاضلی، سری زمانی، یادگیری ماشین

*عهده‌دار مکاتبات:

لادن ریاضی

نشانی: گروه مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، ایران

پست الکترونیکی: l.riazi@gmail.com

۱- مقدمه

در حال حاضر در سرتاسر جهان، حجم بالایی از سرمایه‌ها از طریق بازار سهام معامله می‌شوند و تقریبا اقتصاد ملی هر کشوری با عملکرد بازار سهام آن‌ها در ارتباط است و از آن تاثیر می‌پذیرد. علاوه بر این، اخیرا این گونه بازارها به ابزارهای سرمایه‌گذاری سرمایه‌گذاران حرفه‌ای و مبتدی تبدیل شده است و تمامی اقشار جامعه را به خود جذب می‌کند [1]. به این ترتیب بازار سهام، نقش اساسی در سازماندهی اقتصادی و اجتماعی یک کشور ایفا می‌کند [2]. معاملات بازار سهام، فعالیتی است که در آن سرمایه‌گذاران برای تصمیم‌گیری موثر به اطلاعات سریع و دقیق نیاز دارند. از آنجا که سهام بسیاری در بورس معامله می‌شوند، عوامل متعددی بر روند تصمیم‌گیری تاثیر می‌گذارد. علاوه بر این، رفتار قیمت سهام نامشخص و پیش‌بینی آن دشوار است. به همین دلایل، پیش‌بینی قیمت سهام یک فرایند مهم و چالش‌برانگیز است [3].

سرمایه‌گذاران این بازار همواره می‌خواهند از روند بعدی قیمت‌ها مطلع شوند. به این ترتیب محققان در صدد یافتن روش‌هایی هستند که بتوانند با پیش‌بینی آتی قیمت سهام، سود سرمایه‌گذاران را افزایش دهند. در این راستا پیدا کردن روش‌های مناسب و مبتنی بر اصول علمی در تعیین قیمت آینده سهام برای سرمایه‌گذاران ضروری به نظر می‌رسد [4].

این یک واقعیت انکار ناپذیر است که قیمت سهام به شدت به اخبار، رویدادها و عوامل کلان اقتصادی وابسته است و پیش‌بینی قیمت لحظه‌ای بعدی را به چالش می‌کشد. قیمت تنها بازتابی از احساسات یک بازار یا عملکرد یک شرکت است که به دلیل پیچیدگی متغیرهای اساسی نمی‌توان آن را به‌طور دقیق پیش‌بینی کرد. حتی با الگوریتم‌های یادگیری ماشین، پیش‌بینی دقیق قیمت سهام دشوار است و بهترین چیزی که می‌توان به آن امیدوار بود، شناسایی الگوها یا روندهایی است که ممکن است به تحلیل بازار کمک کند. با این حال، با وجود این محدودیت‌ها، بسیاری از موسسات مالی و محققان از تکنیک‌های یادگیری ماشینی برای پیش‌بینی قیمت سهام استفاده می‌کنند. یکی از دلایل احتمالی این موضوع این است که الگوریتم‌های یادگیری ماشینی می‌توانند الگوهایی را در داده‌های تاریخی شناسایی کنند که ممکن است برای انسان آشکار نباشد. این الگوها می‌توانند برای تجزیه و تحلیل بازار، ارزیابی ریسک و یا حتی تولید سیگنال‌های تجاری مفید باشند. مدل‌های یادگیری ماشین می‌توانند چندین منبع داده از جمله اخبار، رسانه‌های اجتماعی و معیارهای مالی را ترکیب کنند، تا بینشی ایجاد کنند که ممکن است تنها از طریق تجزیه و تحلیل انسانی قابل دستیابی نباشد. به‌طور کلی، در حالی که یادگیری ماشین ممکن است راه‌حل قطعی برای پیش‌بینی قیمت سهام نباشد، اما همچنان می‌تواند ابزار ارزشمندی برای تجزیه و تحلیل رفتار بازار، شناسایی الگوها و بهینه‌سازی عملکرد پرتفوی باشد. با

پذیرش محدودیت‌های یادگیری ماشین و ترکیب آن با تخصص انسانی، می‌توان استراتژی‌های سرمایه‌گذاری مؤثرتری ایجاد کرد.

با ظهور هوش محاسباتی در سه دهه گذشته، مدل‌های غیرخطی بیشتری که با تکنیک‌های شبکه‌های عصبی مصنوعی (ANN)، سیستم‌های عصبی فازی، تکنیک‌های تکاملی و غیره توانمند شده‌اند، توسط بسیاری از محققان برای پیش‌بینی بازار سهام پیشنهاد شده‌اند [5]. بسیاری از تکنیک‌های رایانه‌ای مبتنی بر یادگیری ماشین یا مدل‌های آماری برای برآورد تغییرات قیمت در بازار سهام استفاده شده‌اند [6]. تحقیقات نشان می‌دهد که الگوریتم‌های پیچیده و مدل‌های پیش‌بینی پیچیده و بهینه طراحی‌شده مبتنی بر روش‌های یادگیری ماشین و هوش مصنوعی، افراد را قادر می‌سازد تا قیمت‌های سهام آتی را با درجه بالایی از دقت پیش‌بینی کنند [7]. در این پژوهش برای نخستین بار از تلفیق روش‌های تکامل تفاضلی و جنگل تصادفی برای پیش‌بینی قیمت سهام بهره برده خواهد شد. با تلفیق این روش‌ها می‌توان امید به دستیابی به نتایج موثرتر در زمینه پیش‌بینی قیمت سهام داشت؛ لذا در صدد پاسخگویی به سوالات زیر هستیم:

- چگونه می‌توان با استفاده از روش‌های جنگل تصادفی و روش تکامل تفاضلی راهکار مناسبی برای پیش‌بینی قیمت سهام ارائه داد؟
- تاثیر استفاده از روش تکامل تفاضلی بر افزایش دقت نتایج روش جنگل تصادفی به چه میزان است؟

۲- کارهای مرتبط

قولمز در مقاله‌ی [8]، یک شبکه عمیق LSTM بهینه‌شده با مدل بهینه‌سازی خرگوش‌های مصنوعی (ARO) برای پیش‌بینی قیمت سهام ایجاد کرد. در روش LSTM-ARO الگوریتم ARO، برای بهینه‌سازی فراپارامترهای مدل LSTM و بهبود دقت پیش‌بینی‌های بازار سهام استفاده شد. سهام شاخص DJIA به‌عنوان مجموعه داده استفاده شد که شامل ۳۰ قیمت سهام مختلف است. نتایج نشان داد که LSTM-ARO بر سایر مدل های بررسی‌شده غلبه کرد. ژاو و همکاران در مقاله خود [9] یک مدل رابطه‌ای سری زمانی (TSRM) پیشنهاد کردند که اطلاعات زمان و رابطه را ادغام می‌کند. TSRM از داده‌های معاملات سهام برای به‌دست‌آوردن خودکار طبقه‌بندی سهام از طریق مدل K-means استفاده می‌کند و روابط سهام را استخراج می‌کند. اطلاعات سری زمانی استخراج‌شده با استفاده از حافظه کوتاه‌مدت بلندمدت (LSTM) و یک شبکه کانولوشن گراف (GCN)، برای پیش‌بینی قیمت سهام یکپارچه شد. این روش در بازارهای سهام شانگهای و شنژن چین مورد آزمایش قرار گرفت و نتایج حاکی از بهبود بازده انباشته به ترتیب ۴۴ و ۴۱ درصد بود.

هان و فو در مقاله‌ی خود [10] یک مدل حافظه کوتاه‌مدت دو جهته (Bi-LSTM) برای پیش‌بینی قیمت آتی سهام

بر اساس قیمت‌های تاریخی آن پیشنهاد کردند. مدل Bi-LSTM برای داده‌های تاریخی قیمت سهام شرکت اپل اعمال شد و عملکرد آن با استفاده از میانگین مربعات خطا (MSE) ارزیابی شد. این مدل با MSE 0.00020 در مجموعه تست، عملکرد خوبی از خود نشان داد. در مقاله ژنگ و همکاران [11] عملکرد پیش‌بینی قیمت سهام در زمینه مالی با استفاده از مدل‌های جنگل تصادفی بر روی مجموعه داده‌های اپل، سامسونگ و جنرال الکتریک بررسی شد. طبقه‌بندی‌کننده تصادفی جنگل برای پیش‌بینی روند بلندمدت سهام به دقت قابل‌توجهی دست یافت و توانست به دقت ۸۵ تا ۹۵ درصد برسد. بررسی‌های این تحقیق نشان داد که افزایش تعداد درختان تصمیم در یک جنگل تصادفی می‌تواند به نتایج پایدارتری منجر شود.

ژانگ و همکارانش [12]، یک رویکرد جدید برای پیش‌بینی قیمت سهام با ترکیب رگرسیون بردار پشتیبان (SVR) و الگوریتم کرم شب‌تاب (FA) ارائه می‌دهد. نتایج تجربی نشان داد که در مقایسه با الگوریتم‌های دیگر، الگوریتم MFA پیشنهادشده دارای عملکرد بهتری است. روش پیش‌بینی MFA-SVR می‌تواند به‌عنوان ابزاری موثر برای پیش‌بینی قیمت سهام در نظر گرفته شود. در مقاله‌ی یو و همکارانش [13]، یک الگوریتم کاهش خطی ابعاد برای کاهش بعد عوامل موثر بر قیمت سهام انتخاب شده است. داده‌ها پس از کاهش ابعاد به‌عنوان متغیر ورودی جدید شبکه عصبی بازگشت به عقب برای تحقق پیش‌بینی قیمت سهام استفاده می‌شود. نتایج نشان داد که مدل شبکه عصبی LLE-BP که در این مقاله پیشنهاد شد، دارای دقت پیش‌بینی بالاتری در پیش‌بینی قیمت سهام است. در تحقیق کاو و وانگ [14] برای پیش‌بینی روند آینده فعالیت‌های مالی و شاخص سهام، از شبکه عصبی پیچشی (CNN) استفاده شد. در ابتدا، یک مدل پیش‌بینی شاخص سهام CNN ساخته شد، رابطه پارامترهای ساختاری مدل CNN مورد تجزیه و تحلیل قرار گرفت و یک الگوریتم مدل CNN پیاده‌سازی شد. در ادامه، تاثیر پارامترهای مدل بر نتایج پیش‌بینی مورد بحث قرار گرفت و مدل پیش‌بینی شاخص سهام بر اساس ماشین بردار پشتیبان SVM و CNN ایجاد شد. در نهایت، تجزیه و تحلیل تجربی انجام شد و نتایج نشان داد که هر دو مدل پیش‌بینی امکان‌پذیر و موثر هستند. در مقاله [6]، یک رویکرد مبتنی بر شبکه عصبی عمیق برای پیش‌بینی اینکه آیا قیمت سهام در سال بعد یا در همان سه ماهه ۲۵ درصد افزایش می‌یابد یا خیر، پیشنهاد شد. همچنین روش یادگیری عمیق با رویکردهای جنگل‌های تصادفی و ماشین تقویت گرادیان مقایسه شد. برای آزمایش روش پیشنهادی، پایگاه داده KIS-VALUE متشکل از شاخص قیمت سهام کامپوزیت کره برای دوره ۲۰۰۷ تا ۲۰۱۵ در نظر گرفته شد. روش پیشنهادی عملکرد رضایت‌بخشی را ارائه داد. در تحقیق [7]، طیفی از ده مدل رگرسیون یادگیری عمیق برای پیش‌بینی دقیق قیمت‌های آتی سهام یک شرکت مهم در بخش خودروی هند ارائه شد. نتایج به‌دست‌آمده نشان می‌دهد که مدل‌های CNN سریع‌تر هستند، در حالی که دقت بالاتری توسط مدل‌های LSTM به‌دست می‌آید. در تحقیق [5] بیان شده است که این مطالعه یک مقایسه نرمال‌شده بر روی عملکرد LSTM

و GRU برای پیش‌بینی بازار سهام در شرایط یکسان انجام داد و به‌طور عینی اهمیت ترکیب احساسات اخبار مالی در پیش‌بینی بازار سهام را ارزیابی کرد. آزمایش‌ها نشان می‌دهد که اگر فقط از ویژگی‌های بازار سهام استفاده شود، هر دو LSTM و GRU در پیش‌بینی سهام شرایط یکسانی دارند. عملکرد LSTM و GRU برای پیش‌بینی قیمت سهام را می‌توان با ترکیب احساسات اخبار مالی با ویژگی‌های سهام به‌عنوان ورودی به‌طور قابل‌توجهی بهبود بخشید. در تحقیق [15] در مورد تکنیک‌های رگرسیون در زمینه‌ی پیش بینی قیمت سهام مطالعات جامعی انجام شده است. در این تحقیق برخی از روش‌های رگرسیون مناسب برای پیش‌بینی قیمت در بازار سهام مورد بررسی قرار گرفته است. نتیجه این تجزیه و تحلیل رگرسیون نشان داد که با استفاده از تعداد بیشتری از متغیرها و تکنیک‌های یادگیری ماشین یا تکنیک‌های علم داده نتایج به‌دست‌آمده بهبود قابل‌توجهی خواهد یافت.

در تحقیق بیلاح و همکاران [16]، شبکه‌های حافظه کوتاه‌مدت بلندمدت (LSTM) در زمینه پیش‌بینی سری‌های زمانی تحلیل تجربی شد. این تحقیق مجموعه متنوعی از مجموعه داده‌های دنیای واقعی را تحت تاثیر قرار داد و به میانگین دقت پیش‌بینی ۸۹,۷ درصدی در پیش‌بینی‌های بازار مالی دست یافت که با حاشیه قابل‌توجهی از مدل‌های پایه بهتر بود. در مطالعه ارته و همکارانش [17]، بر روی پیش‌بینی جهت قیمت BTC/USD با استفاده از Random Forest به‌عنوان مدل پیش‌بینی و شاخص‌های فنی و الگوهای کندل استیک به‌عنوان متغیرهای ورودی تمرکز شد. نتایج نشان داد بهترین میانگین سودآوری به‌دست‌آمده ۱,۷۵٪ در هر عملیات بود. در ادامه در جدول ۱ خلاصه‌ی نتایج و مشخصات تعدادی از جدیدترین کارهای انجام شده‌ی گذشته آورده شده است.

جدول ۱: خلاصه‌ی تعدادی از کارهای انجام‌شده‌ی گذشته

تحقیق	روش	مجموعه داده ها	ارزیابی
Gülmez, 2023	LSTM-ARO	سهام AXP سهام AMGN	0.907 R2 0.943 R2
Zhao et al, 2023	TSRM	بازارهای سهام شانگهای بازارهای سهام شنژن چین	بازده انباشته ۴۴ درصد ۴۱ درصد بازده انباشته
Han & Fu, 2023	Bi-LSTM	شرکت اپل	MSE 0.00020
Zheng et al, 2023	Random Forest	مجموعه داده‌های اپل، سامسونگ و جنرال الکتریک دقت ۸۵ تا ۹۵	دقت ۸۵ تا ۹۵ درصد
Billah et al, 2023	LSTM	مجموعه متنوعی از مجموعه داده‌های دنیای واقعی	میانگین دقت ۸۹,۷ درصد
Orte et al, 2023	Random Forest	BTC/USD	میانگین سودآوری ۱,۷۵٪

۳- روش پیشنهادی

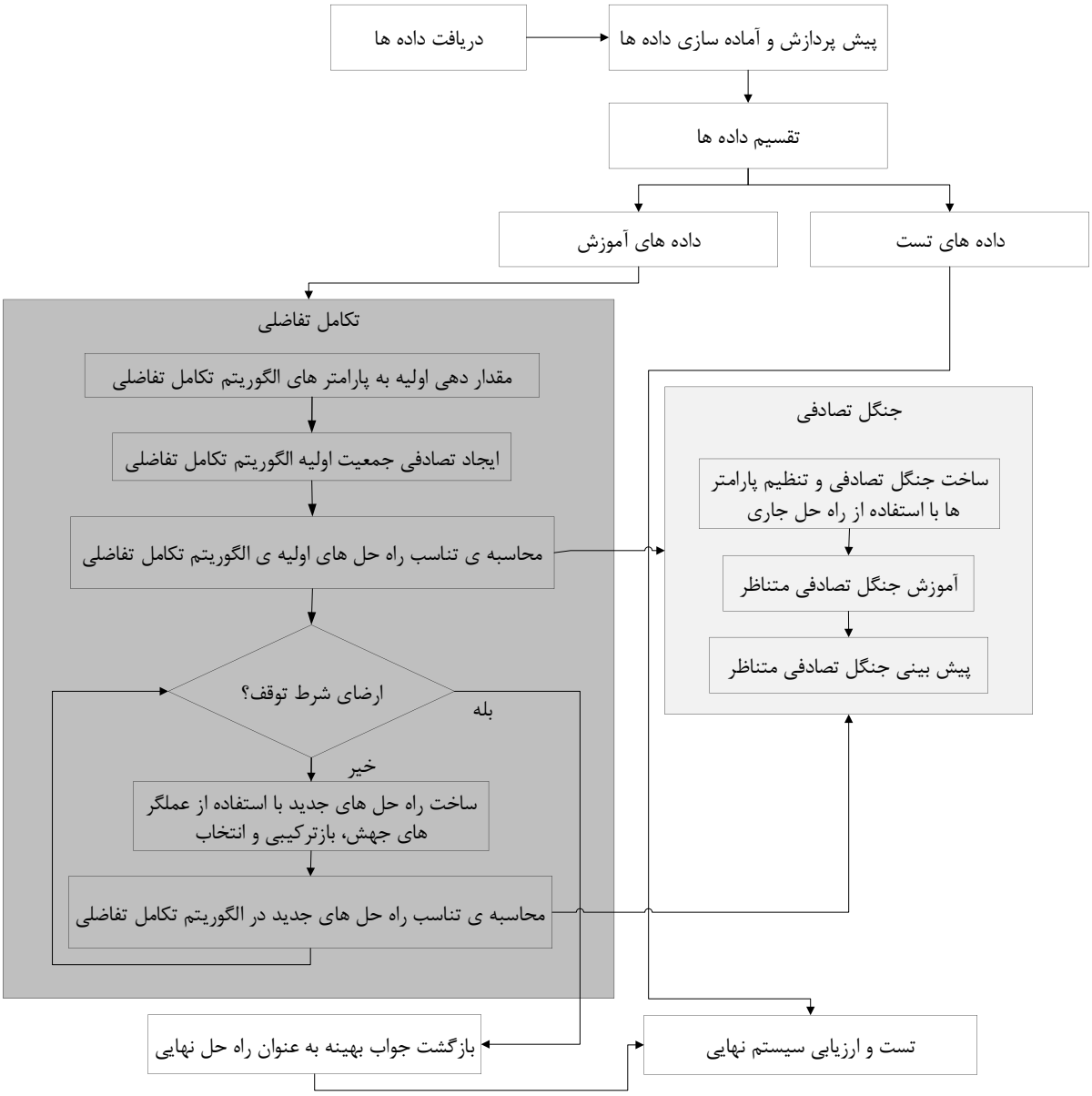
سیستم پیشنهادی این تحقیق برای پیش‌بینی قیمت سهام ترکیب روش‌های جنگل تصادفی و الگوریتم تکامل تفاضلی می باشد. در این تحقیق روش‌های یادگیری ماشین مختلف مورد بررسی قرار گرفت و از بین آن‌ها روش جنگل تصادفی در ترکیب با روش‌های تکاملی به نتایج بهتری دست یافت. جنگل تصادفی برای مدیریت داده‌های با ابعاد بالا که در مجموعه داده های مالی رایج است، مناسب می‌باشد و می‌تواند مقادیر گم‌شده را با انتخاب ویژگی‌های موثر و نمونه‌ها کنترل و روابط غیرخطی بین متغیرها را ثبت کند. جنگل تصادفی نسبت به سایر الگوریتم‌های یادگیری ماشین کمتر مستعد بیش برازش است که باعث می‌شود در برابر نویز و خطا در داده‌های آموزشی مقاوم‌تر باشد.

الگوریتم‌های تکامل تفاضلی در این تحقیق بهترین مقادیر را برای پارامترهای جنگل تصادفی انتخاب می‌کند. وقتی نوبت به آموزش مدل جنگل تصادفی می‌رسد، انتخاب بهترین هایپرپارامترها می‌تواند گامی مهم در دستیابی به عملکرد بهینه باشد. هایپرپارامترها، پارامترهایی هستند که قبل از آموزش مدل تنظیم می‌شوند و می‌توانند به‌طور قابل‌توجهی بر دقت، کارایی محاسباتی و استحکام مدل تاثیر بگذارند. تنظیم هایپرپارامترها، یک مرحله حیاتی در یادگیری ماشین است و انتخاب بهترین هایپرپارامترها می‌تواند فرآیندی زمان‌بر و فشرده باشد. با استفاده از تکنیک‌های بهینه‌سازی مانند تکامل تفاضلی (DE)، می‌توان این فرآیند را خودکار کرد، به‌طوری که بهترین هایپرپارامترها با کارایی بیشتری پیدا شوند. دلیل استفاده از الگوریتم‌های تکامل تفاضلی برای انتخاب پارامترهای موثر این است که این الگوریتم یک الگوریتم بهینه‌سازی محبوب است که به‌طور گسترده در زمینه‌های مختلف از جمله مالی، مهندسی و علوم کامپیوتر استفاده می‌شود. این روش یک نوع الگوریتم بهینه‌سازی مبتنی بر جمعیت است که به‌ویژه در حل مسائل پیچیده بهینه‌سازی موثر است. استحکام، انعطاف‌پذیری، کارایی و سهولت اجرا آن را به انتخابی محبوب تبدیل کرده است. این روش می تواند بهترین هایپرپارامترها را در طیف وسیعی از احتمالات جست‌وجو کند و به سرعت راه‌حل‌های خوبی پیدا کند.

با ترکیب این روش‌ها می‌توان از نقاط قوت هر دو رویکرد استفاده کرد و مدل‌های دقیق و قوی‌تری ایجاد کرد که می‌توانند روابط پیچیده و داده‌های داری نویز را مدیریت کند.

بخش اصلی سیستم پیشنهادی، جنگل تصادفی است که الگوهای موجود در داده‌ها را با دریافت داده‌های آموزشی جست‌وجو می‌کند. پس از پایان مرحله‌ی آموزش، سیستم می‌تواند برای پیش‌بینی قیمت سهام به‌کار گرفته شود. برای افزایش کارایی روش پیشنهادی از الگوریتم تکامل تفاضلی برای بهبود عملکرد جنگل تصادفی استفاده می‌شود. همان‌طور که بیان شد در این سیستم الگوریتم تکامل تفاضلی به دنبال بهینه کردن پارامترهای جنگل تصادفی می‌باشد. به این

ترتیب نتایج به‌دست‌آمده مطلوب‌تر خواهد بود. نحوه‌ی ترکیب الگوریتم تکامل تفاضلی و جنگل تصادفی در شکل ۱ آورده شده است.



شکل ۱: نحوه‌ی ترکیب الگوریتم تکامل تفاضلی و جنگل تصادفی

همان‌طور که در شکل ۱ نشان داده شده است، در سیستم پیشنهادی این تحقیق تعیین مقادیر مناسب برای پارامترهای جنگل تصادفی به وسیله‌ی راه‌حل بهینه‌ی الگوریتم تکامل تفاضلی تعیین می‌شود. در الگوریتم تکامل تفاضلی، هر راه‌حل که در جمعیت اولیه قرار دارد، مقادیر مشخص برای پارامترهای جنگل تصادفی را نشان می‌دهد و در حقیقت معرف

یک جنگل تصادفی خاص می‌باشد. هدف الگوریتم تکامل تفاضلی، پیدا کردن مناسب‌ترین راه‌حل ممکن می‌باشد، به‌طوری که جنگل تصادفی متناظر با آن بهترین عملکرد را داشته باشد.

برای رسیدن به این هدف، در ابتدای کار این الگوریتم، راه‌حل‌های اولیه به‌صورت تصادفی ایجاد می‌شوند و به این ترتیب جمعیت اولیه مقداردهی می‌شود. برای محاسبه‌ی میزان تناسب هر راه‌حل در سیستم پیشنهادی، جنگل تصادفی متناظر با آن راه‌حل ساخته می‌شود و دقت جنگل تصادفی متناظر تعیین می‌گردد. در مرحله‌ی بعد، راه‌حل‌های جدید با اعمال عملیات انتخاب، جهش و بازترکیبی ساخته می‌شود و تناسب آن مشخص خواهد شد و بهترین راه‌حل‌ها به‌عنوان جمعیت در هر تکرار نگه داشته می‌شوند. مراحل ساخت راه‌حل‌های جدید تا زمانی که شرایط توقف الگوریتم خاتمه یابد، تکرار خواهد شد. در نهایت بهترین راه‌حل موجود به‌عنوان نتیجه‌ی نهایی ارائه خواهد شد و جنگل تصادفی متناظر با آن ساخته شده و با داده‌های آموزش، آموزش می‌بیند. در نهایت دقت به‌دست‌آمده از این جنگل تصادفی با داده‌های تست به‌عنوان دقت سیستم پیشنهادی ارائه می‌شود.

۴- معرفی و آماده‌سازی داده‌ها

برای پیاده‌سازی راهکار مورد در این پژوهش، نیاز به مجموعه داده‌های سهام شرکت‌ها می‌باشد. از این رو برای تهیه داده‌های مورد نیاز، از بسترهای آنلاین بهره برده شده است. یکی از این بسترها، وب سایت شناخته‌شده yahoo می‌باشد که در یکی از زیر مجموعه‌های خود، خدمات مالی را نیز ارائه می‌دهد. بخشی از خدمات شامل اطلاعات سهام شرکت‌های مطرح می‌باشد. برای استخراج این داده‌ها، واسط‌های کاربری مناسب موجود می‌باشد که می‌توان از آن‌ها بهره برد. یکی از این واسط‌ها، کتابخانه pandas_datareader می‌باشد که امکان بارگذاری داده‌های سهام شرکت‌های موردنظر را در بازه‌ی زمانی مشخص در محیط پیاده‌سازی پایتون، فراهم می‌نماید.

همان‌طور که مطرح شد، با استفاده از رابط کاربری معرفی‌شده، می‌توان اطلاعات سهام شرکت‌های موردنظر در بازه‌ی زمانی مشخص از بخش مالی وب‌سایت yahoo استخراج نمود. بازه‌ی زمانی مورد بررسی در این پژوهش، محدود به سال‌های ۲۰۱۸ تا ابتدای ۲۰۲۲ می‌باشد. نمایی از اطلاعات استخراج‌شده برای شرکت AMD در جدول ۲ نشان داده شده است.

جدول ۲: نمایی از اطلاعات استخراج‌شده برای شرکت AMD

Date	High	Low	Open	Close	Volume
۰۲-۰۱-۲۰۱۸	۱۱,۰۲۰۰۰۰	۱۰,۳۴۰۰۰۰	۱۰,۴۲۰۰۰۰	۱۰,۹۸۰۰۰۰	۴۴۱۴۶۳۰۰
۰۳-۰۱-۲۰۱۸	۱۲,۱۴۰۰۰۰	۱۱,۳۶۰۰۰۰	۱۱,۶۱۰۰۰۰	۱۱,۵۵۰۰۰۰	۱۵۴۰۶۶۷۰۰

۰۴-۰۱-۲۰۱۸	۱۲,۴۳۰۰۰۰	۱۱,۹۷۰۰۰۰	۱۲,۱۰۰۰۰۰	۱۲,۱۲۰۰۰۰	۱۰۹۵۰۳۰۰۰
۰۵-۰۱-۲۰۱۸	۱۲,۲۲۰۰۰۰	۱۱,۶۶۰۰۰۰	۱۲,۱۹۰۰۰۰	۱۱,۸۸۰۰۰۰	۶۳۸۰۸۹۰۰
۰۸-۰۱-۲۰۱۸	۱۲,۳۰۰۰۰۰	۱۱,۸۵۰۰۰۰	۱۲,۰۱۰۰۰۰	۱۲,۲۸۰۰۰۰	۶۳۳۴۶۰۰۰
...	...	...	...	...	...
۲۳-۱۲-۲۰۲۱	۱۴۹,۰۲۰۰۰۴	۱۴۳,۸۵۰۰۰۶	۱۴۳,۸۸۹۹۹۹	۱۴۶,۱۳۹۹۹۹	۴۸۶۵۳۸۰۰
۲۷-۱۲-۲۰۲۱	۱۵۴,۸۸۹۹۹۹	۱۴۷,۲۵۰۰۰۰	۱۴۷,۵۰۹۹۹۵	۱۵۴,۳۶۰۰۰۱	۵۳۲۹۶۴۰۰
۲۸-۱۲-۲۰۲۱	۱۵۶,۷۲۹۹۹۶	۱۵۱,۳۸۰۰۰۵	۱۵۵,۸۸۰۰۰۵	۱۵۳,۱۴۹۹۹۴	۵۸۶۹۹۱۰۰
۲۹-۱۲-۲۰۲۱	۱۵۴,۳۳۹۹۹۶	۱۴۷,۲۸۹۹۹۳	۱۵۲,۸۲۰۰۰۷	۱۴۸,۲۵۹۹۹۵	۵۱۳۰۰۲۰۰
۲۰۲۱-۱۲-۳۰	۱۴۸,۸۵۰۰۰۶	۱۴۴,۸۵۰۰۰۶	۱۴۷,۴۴۰۰۰۲	۱۴۵,۱۴۹۹۹۴	۴۴۳۵۸۰۰۰

برای استفاده موثرتر از این داده‌ها در یادگیری ماشین، باید فعالیت‌هایی پیرامون آماده‌سازی آن‌ها انجام داد که در ادامه به بیان این فعالیت‌ها می‌پردازیم.

شیفت داده: با توجه به اینکه داده‌های مورد استفاده در این پژوهش در قالب سری زمانی می‌باشند، نیاز به فرآیند شیفت یا جابه‌جایی به میزان روزهای مورد نظر برای پیش‌بینی می‌باشد. این فرآیند شیفت، باید بر روی ویژگی هدف که قصد پیش‌بینی آن را داریم، صورت گیرد. به‌طور معمول در داده‌های سهام، میزان قیمت پایانی یا همان ویژگی close به‌عنوان ویژگی هدف که قصد پیش‌بینی آن را داریم، تعیین می‌شود. بنابراین فرآیند شیفت بر روی ویژگی close و میزان یک واحد صورت می‌پذیرد. با توجه به اینکه میزان شیفت معادل یک واحد تعیین شده است، مدل ساخته‌شده می‌تواند برای تخمین قیمت close برای روز بعد، مورد استفاده قرار گیرد.

نرمال سازی: یکی از فرآیندهای لازم برای آماده‌سازی داده‌ها پیش از مدل‌سازی، نرمال‌سازی داده‌ها می‌باشد. نرمال سازی داده‌ها به تعیین تغییر بازه‌ی تغییرات آن‌ها در بازه‌ای مشخص می‌باشد. این امر سبب می‌شود که مانع از گرایش الگوریتم یادگیری به سمت ویژگی‌هایی با بازه تغییرات زیاد شود.

حذف ویژگی: در صورتی که در داده‌های مورد استفاده، ستون‌های مازادی وجود داشته باشد که کمکی به مدل سازی نمی‌کند، می‌توان آن‌ها را حذف نمود.

استخراج ویژگی‌های تکنیکال: با توجه به اینکه داده‌های استخراج‌شده برای سهام، به‌طور معمول محدود به موارد Open، High، Low، Close و Volume می‌باشد، می‌توان برای غنی‌تر کردن داده‌های مورداستفاده، ویژگی‌های دیگری را بر مبنای همین چند ویژگی پایه استخراج نمود. فرآیند استخراج این ویژگی‌ها پیرامون داده‌های سهام به کمک

تحلیل تکنیکال امکان‌پذیر است. برای این منظور کتابخانه‌های متعددی در زبان پایتون فراهم شده است که در این پژوهش از این کتابخانه‌ها برای استخراج ویژگی‌های تکنیکال استفاده خواهد شد.

بخشی از ویژگی‌های استخراج‌شده برای داده‌های شرکت AMD در جدول ۳ نشان داده شده است.

جدول ۳: بخشی از ویژگی‌های استخراج‌شده برای داده‌های شرکت AMD

Date	High	Low	Open	Close	Volume	Adj Close	volume adi	volume_obv	volume_cmf	...	momentum_wr
۰۲-۰۱-۲۰۱۸	۱۱,۰۲۰۰۰۰	۱۰,۳۴۰۰۰۰	۱۰,۴۲۰۰۰۰	۱۱,۵۵۰۰۰۰	۴۴۱۴۶۳۰۰	۱۰,۹۸۰۰۰۰	+e۱,۱۲۹۶۲۵ ۰۸	۴۴۱۴۶۳۰۰	۲,۵۵۸۸۲۲	...	۷۷,۹۴۱۱۰۲
۰۳-۰۱-۲۰۱۸	۱۲,۱۴۰۰۰۰	۱۱,۳۶۰۰۰۰	۱۱,۶۱۰۰۰۰	۱۲,۱۲۰۰۰۰	۱۵۴۰۶۶۷۰۰	۱۱,۵۵۰۰۰۰	+e۲,۵۹۱۲۸۲ ۰۸	۱۹۸۲۱۳۰۰۰	۱,۳۰۷۳۲۲	...	-۱,۱۱۱۱۳۶
۰۴-۰۱-۲۰۱۸	۱۲,۴۳۰۰۰۰	۱۱,۹۷۰۰۰۰	۱۲,۱۰۰۰۰۰	۱۱,۸۸۰۰۰۰	۱۰۹۵۰۳۰۰۰	۱۲,۱۲۰۰۰۰	+e۱,۰۶۷۷۶۱ ۸	۸۸۷۱۰۰۰۰	۰,۳۴۶۹۹۶	...	-۲۶,۳۱۵۷۹۷
۰۵-۰۱-۲۰۱۸	۱۲,۲۲۰۰۰۰	۱۱,۶۶۰۰۰۰	۱۲,۱۹۰۰۰۰	۱۲,۲۸۰۰۰۰	۶۳۸۰۸۹۰۰	۱۱,۸۸۰۰۰۰	+e۱,۸۴۲۵۸۲ ۰۸	۱۵۲۵۱۸۹۰۰	۰,۴۹۵۹۵۱	...	-۷,۱۷۷۰۶۰
...	...	...	...	...	...	...	...	...	...	...	...
۲۸-۱۲-۲۰۲۱	۱۵۶,۷۲۹۹۹۶	۱۵۱,۳۸۰۰۰۵	۱۵۵,۸۸۰۰۰۵	۱۴۸,۲۵۹۹۹۵	۵۸۶۹۹۱۰۰	۱۵۳,۱۴۹۹۹۴	+e۱,۱۸۶۵۰۸ ۱۰	۲۰۳۰۳۹۱۵۰۰	-۰,۱۱۱۸۰۹	...	-۳۲,۴۱۴۸۶۶
۲۹-۱۲-۲۰۲۱	۱۵۴,۳۳۹۹۹۶	۱۴۷,۲۸۹۹۹۳	۱۵۲,۸۲۰۰۰۷	۱۴۵,۱۴۹۹۹۴	۵۱۳۰۰۲۰۰	۱۴۸,۲۵۹۹۹۵	+e۱۱,۱۷۸۲۶۳ ۱۰	۱۹۷۹۰۹۱۳۰۰	-۰,۰۱۶۹۷۱	...	-۴۴,۳۱۶۹۰۲

تقسیم داده‌ها: داده‌های مورد استفاده باید به مجموعه‌های مجزا برای آموزش و آزمون تقسیم شوند. داده‌های آموزش برای فرآیند آموزش الگوریتم، مورد استفاده قرار می‌گیرد و داده‌های آزمون برای ارزیابی دقت الگوریتم بعد از آموزش مورد استفاده قرار می‌گیرد. تعیین نسبت این تقسیم می‌تواند متفاوت باشد، ولی معمولا قسمت بیشتر داده‌ها برای آموزش و بخش باقی‌مانده برای آزمون لحاظ می‌شود.

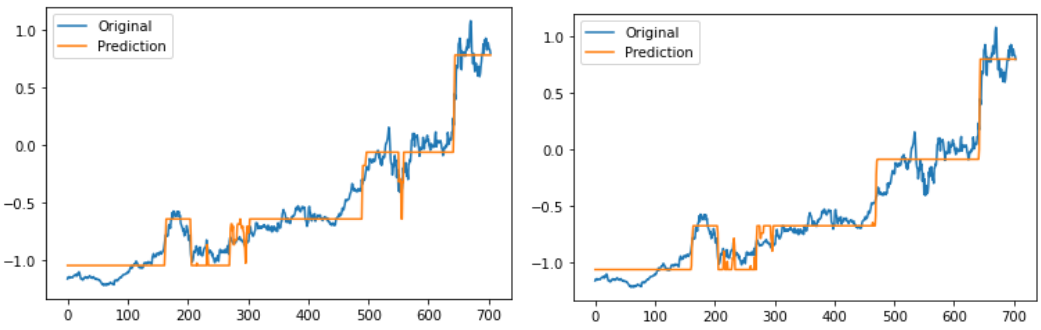
۵- نتایج پیاده‌سازی

برای پیاده‌سازی راهکار موردنظر در این پژوهش از زبان برنامه‌نویسی پایتون و کتابخانه‌های هوش مصنوعی مطرح در این زبان بهره برده شده است. در ادامه‌ی این بخش نتایج به‌دست‌آمده از پیاده‌سازی سیستم پیشنهادی این تحقیق و روش‌های پایه آورده شده است.

نتایج جنگل تصادفی: در این بخش به بررسی اعمال روش جنگل تصادفی بر روی داده‌های استخراج‌شده می‌پردازیم. روش جنگل تصادفی دارای پارامترهای متعددی می‌باشد که در این قسمت از حداقل پارامترها برای ساخت مدل یادگیری بهره برده خواهد شد. در بخش بعدی این پژوهش به بهبود این پارامترها پرداخته خواهد شد.

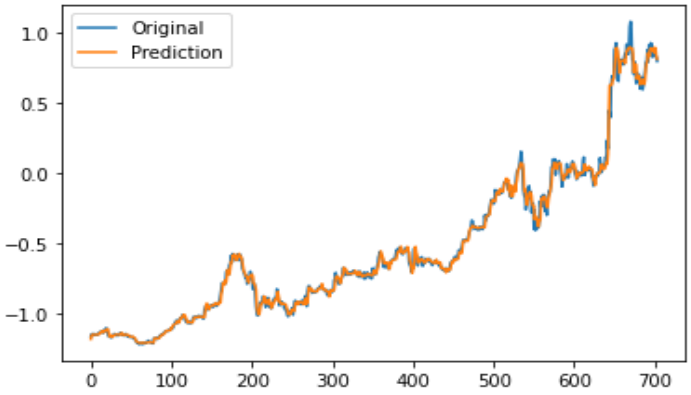
در صورتی که پارامترهای روش جنگل تصادفی به‌صورت نامناسب تنظیم شوند، این روش قادر به پیش‌بینی قیمت سهام به شکل معقول نمی‌باشد. به‌عنوان مثال، شکل ۲ نمونه هایی از مواردی است که روش جنگل تصادفی نتوانسته است پیش‌بینی مناسبی ارائه دهد. خط آبی در شکل‌های مذکور، بیانگر روند تغییرات قیمت سهام شرکت AMD می‌باشد و خط نارنجی بیانگر پیش‌بینی انجام شده توسط روش جنگل تصادفی می‌باشد.

.



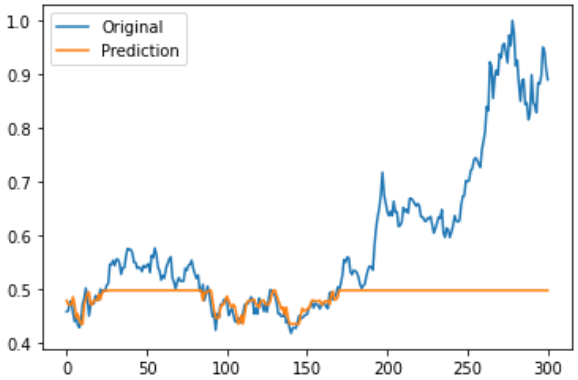
شکل ۲: پیش‌بینی نامناسب قیمت سهام شرکت AMD توسط روش جنگل تصادفی

در صورت تنظیم مناسب پارامترهای شاهد نتایج بهتری خواهیم بود. در شکل ۳، با در نظر گرفتن پارامترهای پیش فرض برای روش جنگل تصادفی و استفاده از کل مجموعه ویژگی‌های آماده‌شده، به میزان خطای ۰/۰۰۱ برای مجموعه داده آموزش دست یافته‌ایم.



شکل ۳: نتیجه حاصل از اعمال پارامترهای پیش‌فرض برای روش جنگل تصادفی و استفاده از کل مجموعه ویژگی‌ها برای پیش‌بینی قیمت سهام شرکت AMD

اما نکته قابل‌توجه در این آزمون این است که علی‌رغم دقت بالا در فرآیند آموزش، دقت در نتایج تست، نامناسب است. شکل ۴ نشان‌دهنده نتیجه اجرا مدل آموزش‌دیده برای داده‌های تست می‌باشد.

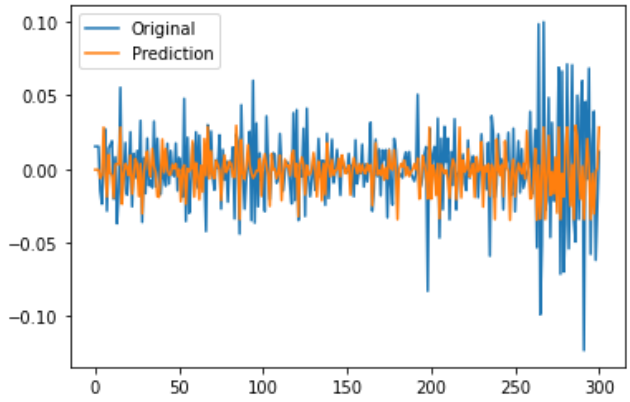


شکل ۴: نتیجه اجرا مدل آموزش‌دیده برای داده‌های تست

علت این نتایج نامناسب، به ماهیت داده‌های مورد بررسی مربوط می‌شود. در صورتی که داده‌های سری زمانی، در طول زمان حالت ایستا[1] نداشته باشند، نمی‌توان مدل مناسبی برای پیش‌بینی آن‌ها ساخت. منظور از ایستا بودن داده‌های سری زمانی از نظر شهودی این است که در طول زمان دارای روندی نسبتا تکراری باشند و نظر محاسباتی به این معنی است که مفاهیم آماری نظیر میانگین، واریانس و کوواریانس باید در طول زمان، ثابت باشد. در بخشی بعدی، تغییراتی بر روی داده‌ها اعمال می‌شود تا داده‌های مورد بررسی به‌صورت ایستا تبدیل شوند.

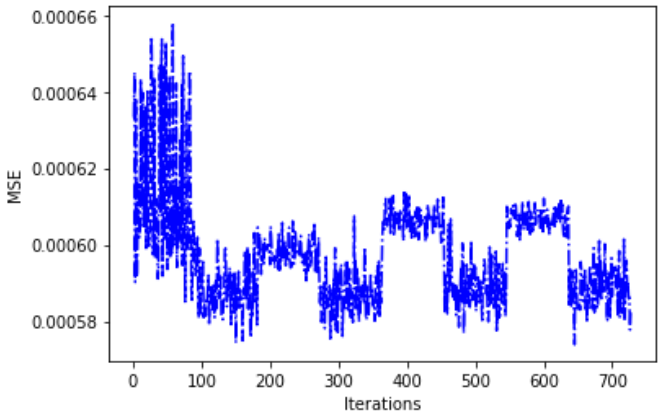
[1] stationary

نتایج جنگل تصادفی برای داده‌های ایستا شده: از بین روش‌های مطرح شده برای ایستاسازی داده‌ها با توجه به این که راه‌حل‌های لگاریتمی و ریشه‌گیری برای اعداد منفی کارایی ندارد؛ در این پژوهش از روش محاسبه اختلاف داده‌ها در طول سری زمانی استفاده شده است. در این روش، به جای استفاده از داده‌های اصلی، از اختلاف مقدار داده‌ها در واحد زمانی جاری و واحد زمانی بعد استفاده می‌شود. شکل ۵ نتیجه حاصل از آزمون با روش جنگل تصادفی برای داده‌های ایستا شده را نشان می‌دهد.



شکل۵: نتیجه حاصل از آزمون با روش جنگل تصادفی برای داده‌های ایستا شده

نتایج سیستم پیشنهادی (ترکیب جنگل تصادفی و تکامل تفاضلی): با توجه به اینکه روش جنگل تصادفی دارای پارامترهای متعددی می‌باشد و هر این از این پارامترها می‌توانند مقادیر متعددی بپذیرند، بنابراین باید راهکاری برای تعیین مناسب این پارامترها برای پیش‌بینی قیمت سهام ارائه نمود. همچنین با توجه به اینکه، تعداد ویژگی‌های مورد بررسی در این پژوهش با توجه به اضافه شدن ویژگی‌های تکنیکال نسبتا زیاد می‌باشد، می‌توان فرآیند انتخاب ویژگی را نیز لحاظ نمود. در این بخش، با بهره‌گیری از روش تکامل تفاضلی، به‌طور همزمان هم ساختار جنگل تصادفی و هم تعیین زیرمجموعه مناسب مورد بررسی قرار می‌گیرد. شکل ۶ نشان‌دهنده روند بهبود میزان خطا در داده‌های تست در طول اجرای روش تکامل تفاضلی می‌باشد.



شکل ۶: روند بهبود میزان خطا در داده‌های تست در طول اجرای روش تکامل تفاضلی

بهترین نتیجه حاصل در فرآیند اجرای روش تکامل تفاضلی، برای داده‌های آموزش دارای میزان خطای ۰/۰۰۰۱۴۷۰۷۱۶ و برای داده‌های تست دارای میزان خطای ۰/۰۰۰۵۷۱۸۵۶ می‌باشد. همچنین پارامترهای روش جنگل تصادفی که منجر به دست یابی به این نتیجه شده است در جدول ۴ آورده شده اند.

جدول ۴: بخشی از ویژگی‌های استخراج‌شده برای داده‌های شرکت AMD

پارامتر	Criterion	max_depth	min_samples_leaf	min_samples_split	n_estimators
پارامتر	تابع عملکرد اندازه گیری کیفیت فرآیند تقسیم برای ساخت جنگل	حداکثر عمق درخت	حداقل تعداد نمونه مورد نیاز برای قرار گرفتن در یک گره برگ	حداقل تعداد نمونه مورد نیاز برای تقسیم یک گره داخلی	تعداد درختان جنگل
مقدار	'mae'	۳	۲	۶	۳۰

همچنین برخی از ویژگی‌های مهم استفاده‌شده برای دستیابی به این نتیجه و مفهوم هر یک از آن‌ها به‌صورت زیر می‌باشد:

- High: بالاترین قیمتی که یک سهم در طول روز معاملاتی با آن معامله شد.
- Open: قیمت در دوره شروع معاملات
- Adj Close: قیمت بسته شدن پس از تعدیل برای همه تقسیمات قابل اجرا و تقسیم سود
- volume_cmf: یک شاخص تحلیل تکنیکال است که برای اندازه‌گیری حجم جریان پول در یک دوره زمانی معین استفاده می‌شود.

- volume_sma_em: میانگین قیمت در دوره مشخص شده
- volatility_atr: میانگین دامنه‌های واقعی در دوره مشخص شده
- momentum_kama: میزان افزایش یا کاهش قیمت سهام را اندازه‌گیری می‌کند

مقایسه‌ی روش پیشنهادی و روش‌های پایه: مقایسه نتایج حاصل از راهکار پیشنهادی و سایر روش‌های پایه از جمله روش جنگل تصادفی، شبکه عصبی و رگرسیون خطی در جدول ۵ نشان داده شده است. با توجه به جدول ۴، به ترتیب روش‌های جنگل‌های تصادفی، شبکه عصبی و رگرسیون خطی به میزان ۷۴/۹٪، ۵۰/۷٪ و ۹۶/۶٪ نسبت به روش پیشنهادی دارای خطای آموزشی و همچنین به ترتیب به میزان ۳۷/۵٪، ۳۰/۴٪ و ۹۸/۹٪ نسبت به روش پیشنهادی دارای خطا تست می باشند.

جدول ۵: مقایسه نتایج حاصل از راهکار پیشنهادی و سایر روش‌ها

	میزان خطای آموزش	درصد میزان خطای آموزش	میزان خطای تست	درصد میزان خطای تست
جنگل تصادفی (پارامترهای پیش‌فرض)	۰,۰۰۰۵۸۶۱	۷۴,۹	۰,۰۰۰۹۱۵	۳۷,۵
شبکه عصبی	۰,۰۰۰۲۹۸۷	۵۰,۷	۰,۰۰۰۸۲۱	۳۰,۴
رگرسیون خطی	۰,۰۰۴۳۵۲۳	۹۶,۶	۰,۰۰۵۲۳۱۶	۹۸,۹
روش پیشنهادی	۰,۰۰۰۱۴۷۰	-	۰,۰۰۰۵۷۱	-

۶- نتیجه‌گیری

پیش‌بینی قیمت سهام به دلیل پیچیدگی و عدم قطعیت آن به یک موضوع بسیار چالش‌برانگیز تبدیل شده است. برای رفع این مشکل که پیامدهای اقتصادی قابل‌توجهی دارد، تکنیک‌های مختلفی توسط محققان توسعه داده شده است. رویکردهای مبتنی بر یادگیری ماشین به‌عنوان یکی از امیدوارکننده‌ترین تکنیک‌ها در این زمینه ظاهر شده‌اند. برای پیش‌بینی بهتر قیمت سهام و سپس ارائه پیشنهادهای معقول به سهامداران، این تحقیق یک مدل ترکیبی مبتنی بر تکنیک‌های جنگل تصادفی و روش تکامل تفاضلی پیشنهاد کرده است. برای به‌دست‌آوردن نتایج مناسب در این تحقیق ایستاسازی داده‌ها انجام شده است. برای انجام این کار به جای استفاده از داده‌های اصلی، از اختلاف مقدار داده‌ها در واحد زمانی جاری و واحد زمانی بعد استفاده شده است که تاثیر قابل‌توجهی در نتایج داشته است. مدل مورد استفاده با استفاده از روش تکامل تفاضلی توانسته پارامترهای جنگل تصافی را به خوبی تنظیم کند. نتایج پیاده‌سازی سیستم

پیشنهادی و مقایسه‌ی آن با روند تغییرات قیمت سهام دقت مناسب سیستم پیشنهادی را نشان می‌دهد. در ادامه به بررسی پرسش‌های پژوهش پرداخته می‌شود:

۱-چگونه می‌توان با استفاده از روش‌های جنگل تصادفی و روش تکامل تفاضلی راهکار مناسبی برای پیش‌بینی قیمت سهام ارائه داد؟

با فرض آن که بین داده‌های جمع‌آوری‌شده برای پیش‌بینی قیمت سهام رابطه‌هایی وجود دارد که قابل استخراج است و با در نظر گرفتن پارامترهای پیش فرض برای روش جنگل تصادفی و استفاده از کل مجموعه ویژگی‌های آماده شده، به میزان خطای ۰/۰۰۱ برای مجموعه داده آموزش دست یافته‌ایم.

بهترین نتیجه حاصل در فرآیند اجرای روش تکامل تفاضلی، برای داده‌های آموزش دارای میزان خطای ۰/۰۰۰۱۴۷۰۷۱۶ و برای داده‌های آزمون دارای میزان خطای ۰/۰۰۰۵۷۱۸۵۶ می‌باشد. همچنین با توجه به اینکه روش جنگل تصادفی دارای پارامترهای متعددی می‌باشد و هر این از این پارامترها می‌توانند مقادیر متعددی بپذیرند، بنابراین باید راهکاری برای تعیین مناسب این پارامترها برای پیش‌بینی قیمت سهام ارائه نمود که با تنظیم مناسب پارامترها و ایستاسازی داده‌ها شاهد نتایج بهتری بودیم.

۲-تاثیر استفاده از روش تکامل تفاضلی بر افزایش دقت نتایج روش جنگل تصادفی به چه میزان است؟

با فرض اینکه اعمال روش تلفیقی مبتنی بر جنگل تصادفی به تکامل تفاضلی می‌تواند منجر به بهبود نتایج در زمینه پیش‌بینی قیمت سهام شود. مدل مورد استفاده در این تحقیق با استفاده از روش تکامل تفاضلی توانسته پارامترهای جنگل تصافی را به درستی تنظیم کند و نتایج حاصل از این تحقیق بیانگر این است که استفاده از تلفیق روش‌های جنگل تصادفی و روش تکامل تفاضلی به این صورت که روش تکامل تفاضلی برای فرآیند انتخاب ویژگی و بهینه‌سازی پارامترها مورد استفاده قرار بگیرد و می‌تواند به نتایج مناسبی در حوزه پژوهش دست یافت. همچنین با توجه به اینکه، تعداد ویژگی‌های مورد بررسی در این پژوهش با توجه به اضافه شدن ویژگی‌های تکنیکال، نسبتا زیاد می‌باشد، می‌توان فرآیند انتخاب ویژگی را نیز لحاظ نمود. بهترین تاثیر و نتیجه حاصل در فرآیند اجرای روش تکامل تفاضلی، برای داده‌های آموزش دارای میزان خطای ۰/۰۰۰۱۴۷۰۷۱۶ و برای داده‌های آزمون دارای میزان خطای ۰/۰۰۰۵۷۱۸۵۶ می‌باشد.

به‌طور کلی نتایج حاصل از پیاده‌سازی روش پیشنهادی، بیانگر برتری و بهبود روش مورد نظر نسبت به راهکارهایی نظیر، جنگل تصادفی، شبکه عصبی و رگرسیون خطی بوده است.

شایان ذکر است در این پژوهش به پیش‌بینی قیمت سهام به کمک تلفیق روش‌های جنگل تصادفی و الگوریتم تکامل تفاضلی پرداخته شد. برای ادامه این پژوهش و بررسی راهکارهای بیشتر می‌توان از سایر روش‌های تکاملی برای بهبود ساختار روش جنگل تصادفی و انتخاب زیرمجموعه داده‌ها پرداخت. همچنین به جای استفاده از راهکار جنگل تصادفی، می‌توان از سایر راهکارهای یادگیری ماشین و یادگیری عمیق بهره برد و کارایی آن‌ها را مورد بررسی قرار داد. از جمله این راهکارهای یادگیری ماشین می‌توان به شبکه عصبی، درخت تصمیم، نایو بیز و ماشین بردار پشتیبان اشاره کرد. همچنین تلفیق نتایج این روش‌ها با استفاده از الگوریتم‌هایی نظیر رای‌گیری و پشته‌سازی نیز می‌تواند مورد توجه پژوهشگران قرار گیرد.

با توجه به در دسترس بودن داده‌های مورد نیاز برای پیاده‌سازی، به‌طور مشخص داده‌های سهام شرکت ها، محدودیت به خصوصی برای این پژوهش وجود نداشت. البته برای گسترش نتایج این پژوهش برای داده‌های سهام بازار بورس تهران، به وجود API لازم برای دسترسی به داده‌های سهام شرکت‌های داخلی در بازار بورس نیاز است.

مراجع

[1]. B.M. Blau, "Income inequality, poverty, and the liquidity of stock markets," *Journal of Development Economics*, vol. 130, pp.113–126, 2017.

[2]. G. Kumar, J. Sanjeev, S.P. Uday, "Stock market forecasting using computational intelligence: A survey," *Archives of Computational Methods in Engineering*, 1-33, 2020.

[3]. M. Obthong, T. Nongnuch J. Watthanasak, W. Gary, "A survey on machine learning for stock price prediction: algorithms and techniques," 63-71, 2020.

[4]. N. Forouzanmanesh, *Stock price prediction using expert opinion and fuzzy-neural system*, Master's thesis, University of Qom, Seattle, 2016. [Persian]

[5]. T.B. Shahi, S. Ashish, N. Arjun, W. Guo, "Stock price forecasting with deep learning: A comparative study," *Mathematics*, vol. 8, no. 9, p. 1441, 2020.

[6]. S.S. Roy, C. Rohan, C.L. Kun, S. Concetto, M. Behnam, "Random Forest, gradient boosted machines and deep neural network for stock price forecasting: a comparative analysis on South Korean companies," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 33, no. 1, 62-71, 2020.

[7]. J. Sen, M. Sidra, "Accurate stock price forecasting using robust and optimized deep learning models, " arXiv preprint arXiv:2103.15096, 2021.

[8]. B. Gülmez, “Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm,”.Expert Systems with Applications, 227, p.120346, 2023.

[9]. C. Zhao, P. Hu, X. Liu, X. Lan, H. Zhang, “Stock market analysis using time series relational models for stock price prediction,” Mathematics, 24;11(5):1130, 2023.
[10]. C. Han, X. Fu, “Challenge and opportunity: deep learning-based stock price prediction by using Bi-directional LSTM model,” Frontiers in Business, Economics and Management, 2;8(2):51-4, 2023.
[11]. J. Zheng, D. Xin, Q. Cheng, M. Tian, L. Yang, “The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance,” arXiv preprint arXiv:17194, 2024.
[12]. J. Zhang, T. Yu-Fan, C. Wei, "Support vector regression with modified firefly algorithm for stock price forecasting," *Applied Intelligence*, vol. 49, no. 5, pp. 1658-1674, 2019.
[13]. Z. Yu, Q. Lu, C. Yunjing, D.P. Milan, "Stock price forecasting based on LLE-BP neural network model," *Physica A: Statistical Mechanics and its Applications*, vol. 553, p. 124197, 2020.
[14]. J. Cao, J. Wang, "Stock price forecasting model based on modified convolution neural network and financial time series analysis," *International Journal of Communication Systems*, vol. 32, no. 12, p. e3987, 2019.
[15]. B.K. Gupta, K.M. Manas, H. Sarbeswara, "Survey on Stock Price Forecasting Using Regression Analysis," *In* Intelligent and Cloud Computing, pp. 147-156, 2021.
[16]. M.M, Billah, A. Sultana, F. Bhuiyan, M.G. Kaosar, “Stock price prediction: comparison of different moving average techniques using deep learning model,” Neural Computing and Applications, 36(11), pp.5861-5871, 2024.
F. Orte, J. Mira, M.J. Sánchez, P. Solana, “A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction,” Research in International Business and Finance. 1;64:101829, 2023