

انتخاب بهینه سبد سهام با استفاده از الگوریتم ترکیبی فرا ابتکاری

محمدباقر یزدانی خداهری^۱، سیدحسین نسل موسوی^{۲*}، میرسعید حسینی شیروانی^۳

۱- دانشجوی دکتری، گروه حسابداری، واحد قائم شهر، دانشگاه آزاد اسلامی، قائم شهر، ایران

۲- استادیار، گروه حسابداری، واحد قائم شهر، دانشگاه آزاد اسلامی، قائم شهر، ایران

۳- استادیار، گروه کامپیوتر، واحد ساری، دانشگاه آزاد اسلامی، ساری، ایران

رسید مقاله: ۵ بهمن ۱۳۹۸

پذیرش مقاله: ۲ شهریور ۱۳۹۹

چکیده

انتخاب سبد سهام همواره از اساسی ترین مسایل سرمایه گذاران است. از لحاظ نظری انتخاب سبد سهام در حالت حداقل کردن ریسک با فرض ثبات بازده به کمک روابط ریاضی قابل حل است، اما با تنوع انتخاب در بازار سرمایه، تنها روابط ریاضی، راه حل موثری نیست. تنوع ابزارهای سرمایه گذاری و تفاوت تابع مطلوبیت سرمایه گذاران، فرآیند انتخاب را پیچیده نموده است. اینک گسترش بازارهای مالی و سرمایه، استفاده از سیستم های قاعده محور برای تصمیم گیری های سریع، با ریسک حداقل و به دور از اشتباهات انسانی، طراحی، توسعه یا بهبود این سیستم ها می تواند به عنوان مزیت رقابتی باشد. در پژوهش حاضر، از الگوریتم شبکه های عصبی و برنامه نویسی شبکه ژنتیک، برای تشخیص ویژگی های موثر و از درخت تصمیم ID3 بهبود یافته به عنوان روش پیشنهادی جهت پیش بینی قیمت و روند تغییر قیمت سهام برای انتخاب سبد بهینه استفاده شده است. نتایج تحقیق نشان می دهد، روش پیشنهادی علاوه بر کاهش سربار محاسباتی و حافظه ای، قادر است تا با دقت بالایی به پیش بینی نوسانات شدید با الگوهای غیرخطی پرداخته و نسبت به روش های روز دنیا چون جستجوی نزدیک ترین همسایگی، رگرسیون خطی، میانگین متحرک خود همبسته و الگوریتم پیش بینی سری زمانی، بهتر عمل نماید.

کلمات کلیدی: سبد بهینه سهام، درخت تصمیم بهبود یافته، الگوریتم ژنتیک، شبکه های عصبی.

۱ مقدمه

یکی از اهداف اصلی گردانندگان بازارهای پولی و مالی این است که هر کسی با هر سلیقه و هر مقدار و با انتخاب هر نوع دارایی بتواند وارد این بازارها شده و فرصت های مناسب سرمایه گذاری را تشخیص دهد و سود مناسبی کسب نماید. در همین راستا، از مهم ترین رسالت مهندسی مالی، طراحی ابزارهای متنوع مالی به منظور تنوع بخشیدن به بازارهای پولی و مالی می باشد [۱-۳].

* عهده دار مکاتبات

آدرس الکترونیکی: Nseyedhossein2@gmail.com

تحقیقات [۳،۴،۱] نشان می‌دهند که طراحی و ارایه مدل‌های قاعده محور برای زمان خرید و فروش امکان پذیر بوده و می‌توان با تولید قواعد معاملاتی، سیستم‌های توانمندی را برای پشتیبانی از تصمیم‌گیری‌های سرمایه‌گذاران، توسعه داد. انتخاب سبد سرمایه با تخصیص سرمایه محدود به تعدادی از دارایی‌های بالقوه سرمایه‌گذاری به منظور دستیابی به استراتژی سرمایه‌گذاری سودآور، در ارتباط است [۵]. سبد سرمایه، ترکیبی مناسب از سهام یا سایر دارایی‌ها است که یک سرمایه‌گذار آن‌ها را خریداری می‌کند. هدف از تشکیل سبد سرمایه، توزیع ریسک سرمایه‌گذاری بین چند سهم است، به گونه‌ای که سود یک سهم بتواند ضرر سهام دیگر را جبران کند [۴،۱]. در مساله بهینه‌سازی سبد سرمایه، هدف کمینه‌سازی ریسک و بیشینه‌سازی سود است. بنابراین ایجاد توازن بین ریسک و بازده، امری ضروری است [۲،۴،۵،۶].

ایجاد این توازن به نظر ساده است ولیکن در عمل روش‌های مختلفی برای تشکیل سبد سرمایه‌گذاری، استفاده می‌شود. هدف از حل مدل‌های انتخاب سبد سرمایه‌گذاری، ارایه مجموعه‌ای از جواب‌ها جهت انتخاب سبد سرمایه مورد نظر به تصمیم‌گیرندگان است.

دستیابی به رشد اقتصادی و ایجاد انگیزه برای سرمایه‌گذاری، زمانی تسریع می‌گردد که بازارهای سرمایه فعال و قابل اعتماد باشند. وجود بازارهای بورس فعال همواره سرمایه‌گذاران متعددی را به تکاپو داشته و حرکت جریان سرمایه و منابع مالی را به بخش‌های مولد تسریع می‌نماید. کلید موفقیت سرمایه‌گذاران در بازار سرمایه، اتخاذ رویکردی مناسب برای مدیریت سهام ساخته شده توسط آنها است [۳،۱].

اولین مطالعه در رابطه با مساله انتخاب پورتفولیو، به کارگیری مفهوم مجموعه کارا است که توسط مارکوویتز (۱۹۵۲) ارایه شد [۴]. مطالعه او که مبنای تئوری پورتفولیوی مدرن به شمار می‌رود، انتخاب پورتفولیو را در قالب یک مساله بهینه‌سازی میانگین-واریانس با دو معیار اساسی، حداکثرسازی سود (میانگین بازده مورد انتظار) و حداقل‌سازی ریسک (واریانس بازده مورد انتظار) مورد توجه قرار داده، لذا یک پورتفولیوی مطلوب با توجه به بده بستان ریسک و بازده مورد انتظار تعیین می‌شود [۷،۱].

مدل مارکوویتز، اگرچه مبنای تئوری پورتفولیوی مدرن است، به دلیل اینکه مدلی ساده‌سازی شده با مفروضات غیرواقع‌گرایانه است، استفاده عملی چندانی ندارد. از نقطه نظر عملی، سرمایه‌گذاران معمولاً با محدودیت‌های الزام‌آوری روبه‌رو هستند. هر چقدر محدودیت‌های عملی به منظور توسعه مدل در مدل‌سازی افزایش یابد، حل مدل مشکل‌تر می‌شود. بسیاری از محققان همچون شریعت پناهی و همکارانش [۴] انتخاب پورتفولیو را در قالب یک مساله بهینه‌سازی میانگین-واریانس استفاده نموده‌اند. محققان از انواع تکنیک‌ها برای حل مساله انتخاب پورتفولیوی مقید استفاده کرده‌اند [۸]. هدف از حل مدل‌های انتخاب سبد سرمایه‌گذاری، ارایه مجموعه‌ای جواب‌ها به تصمیم‌گیرندگان جهت انتخاب سبد سرمایه مورد نظر است [۹].

در این تحقیق با توجه به اینکه روش‌های فراابتکاری هر یک دارای معایبی می‌باشند و مزایایی نیز در کنار این معایب دارا می‌باشند، سعی شده است تا راهکارهای فراابتکاری قابل ادغام انتخاب شوند و به نوعی از آن‌ها استفاده شود که در کنار شامل شدن تمامی مزایای این روش‌ها، دارای هیچ یک از معایب

این روش‌ها نباشد و از این حیث دارای برتری نسبت به استفاده از هر یک از این روش‌ها به تنهایی باشد. در این پژوهش سوال اساسی‌ای که وجود دارد، آن است که چگونه می‌توان با استفاده از راهکاری ترکیبی، تغییرات قیمت سهام را پیش‌بینی نمود و با استفاده از آن سبد بهینه سهام را شناسایی کرد؟ هدف از تحقیق حاضر، استفاده از الگوریتم‌های فراابتکاری در جهت ارایه راهکاری است تا بتوان به‌وسیله آن میزان تغییرات قیمت هر یک از سهام را پیش‌بینی و سهامی که دارای بیش‌ترین تغییرات از لحاظ افزایش قیمت هستند را شناسایی و در قالب سبد سهام ارایه و بدین شکل بتوان بهترین سبد سهام را یافت و به تبع آن بیش‌ترین سود را نیز کسب نمود. برای پیش‌بینی در این راهکار از درخت تصمیم بهبود داده شده‌ای که دارای کم‌ترین ارتفاع و پایین‌ترین سربار محاسباتی باشد، استفاده شده است.

۲ مبانی نظری و ادبیات پژوهش

به‌طور کلی الگوریتم‌هایی که برای حل مسایل بهینه‌یابی وجود دارند، شامل الگوریتم‌های دقیق و الگوریتم‌های تقریبی هستند. الگوریتم‌های دقیق، اگرچه قادر به یافتن جواب بهینه به صورت دقیق هستند، اما در مورد مسایل بهینه‌سازی سخت، کارایی کافی ندارند و زمان اجرای آنها متناسب با ابعاد مسایل به صورت نمایی افزایش می‌یابد. الگوریتم‌های تقریبی، قادر به یافتن جواب‌های نزدیک به بهینه در زمان حل کوتاه برای مسایل بهینه‌سازی سخت هستند. این الگوریتم‌ها به سه دسته، ابتکاری^۱، فرا ابتکاری^۲، فوق ابتکاری^۳ تقسیم می‌شوند [۱۰].

دو مشکل اصلی الگوریتم‌های ابتکاری، گیر افتادن آنها در نقاط بهینه محلی، همگرایی زودرس به این نقاط است. الگوریتم‌های فرا ابتکاری برای حل مشکلات الگوریتم‌های ابتکاری ارایه شده و در واقع الگوریتم‌های فراابتکاری، یکی از انواع الگوریتم‌های بهینه‌سازی تقریبی هستند که دارای راهکارهای برون رفت از نقاط بهینه محلی هستند و قابلیت کاربرد در طیف گسترده‌ای از مسایل را دارند [۱۱، ۱۰].

الگوریتم‌های شناخته شده فراابتکاری بر پایه جمعیت، عبارتند از الگوریتم‌های تکاملی ژنتیک، برنامه‌ریزی ژنتیک، بهینه‌سازی کلونی مورچگان، زنبورها، روش بهینه‌سازی ازدحام ذرات، بهینه‌سازی ملهم از فیزیک نور، الگوریتم ریشه-پاجوش و از معروف‌ترین الگوریتم‌های فراابتکاری جدیدی که با توجه به موجودات زنده موجود در طبیعت، اخیراً توسعه یافته‌اند، الگوریتم بهینه‌سازی گرگ خاکستری، سنجاقک، گرده افشانی گل‌ها، نهنگ یا وال و ملخ می‌باشند [۱۲].

۲-۱ مفهوم سرمایه‌گذاری و بورس

به‌طور کلی سرمایه‌گذاری عبارت است از تبدیل وجوه مالی به یک یا چند نوع دارایی دیگر و نگهداری آن برای مدتی در زمان آینده، به همراه پذیرفتن ریسک مشخص یا نامشخص، برای کسب سود در آینده است. به مجموعه‌هایی از رفتارها، فرآیندها و قوانین که سرمایه‌گذار را به سمت بهترین سرمایه‌گذاری هدایت می‌کند،

1. heuristic
2. meta-heuristic
3. hyper heuristic

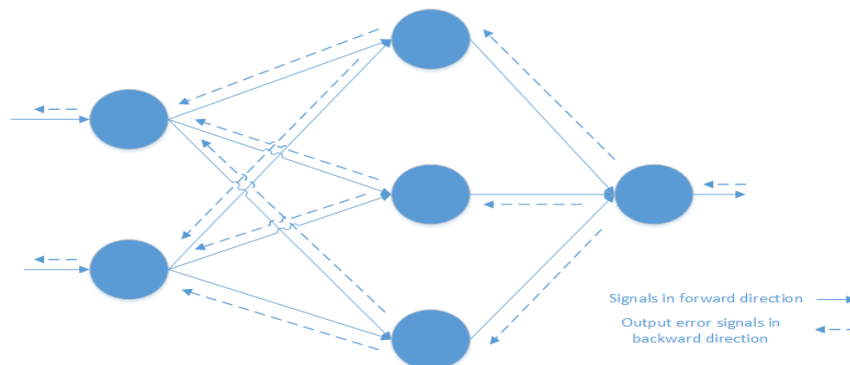
استراتژی‌های سرمایه‌گذاری می‌گویند [۲]. عوامل مختلفی مانند اهداف بلند مدت و کوتاه مدت، آستانه ریسک‌پذیری و اهداف شخصی سرمایه‌گذار در به‌کارگیری این استراتژی‌ها نقش دارند [۹، ۴].

۲-۲ الگوریتم ژنتیک

الگوریتم ژنتیک، یکی از زیرمجموعه‌های محاسبات تکامل یافته است و رابطه مستقیمی با مبحث هوش مصنوعی دارد. در حقیقت الگوریتم ژنتیک، به‌عنوان زیرمجموعه هوش مصنوعی، یک روش جستجوی کلی است که از قوانین تکامل بیولوژیک طبیعی تقلید می‌کند. این الگوریتم بر روی یک سری از جواب‌های مساله، به امید به دست آوردن جواب‌های بهتر، قانون بقای بهترین را اعمال می‌کند. در هر نسل به کمک فرآیند انتخابی متناسب با ارزش جواب‌ها و تولید مثل جواب‌های انتخاب شده به کمک عملگرهایی که از ژنتیک طبیعی تقلید شده‌اند، تقریب‌های بهتری از جواب نهایی به دست می‌آید [۱۳]. این فرایند باعث می‌شود که نسل‌های جدید با شرایط مساله سازگارتر باشد.

۲-۳ شبکه پرسپترون چند لایه

شبکه پرسپترون چند لایه، عملاً از به هم پیوستن سه شبکه پرسپترون تک لایه ایجاد شده و اتصال فقط از جزء i به جزء $i + 1$ برقرار است و در جهت عکس صادق نیست. این شبکه دارای یک لایه خروجی و دو لایه میانی است. خروجی‌های لایه اول، بردار ورودی لایه دوم را تشکیل داده و بردار خروجی لایه دوم، ورودی‌های لایه سوم را می‌سازند و خروجی‌های لایه سوم، پاسخ واقعی شبکه را تشکیل می‌دهند. به‌عبارتی روشن‌تر، روند جریان سیگنالی در شبکه، در یک مسیر پیش‌خور صورت می‌گیرد (از چپ به راست از لایه‌ای به لایه دیگر). هر لایه می‌تواند از تعدادی نرون مختلف با توابع تبدیل متفاوت، برخوردار باشد. یعنی مدل‌های نرون‌ها در لایه‌ها می‌توانند متفاوت در نظر گرفته شوند. در طی آموزش شبکه MLP به کمک الگوریتم یادگیری BP، ابتدا محاسبات از ورودی شبکه به سوی خروجی شبکه انجام و سپس مقادیر خطای محاسبه شده به لایه‌های قبل انتشار می‌یابد [۱۴]. به حرکت سیگنال‌ها در مسیر رفت (از سمت چپ به راست شبکه)، سیگنال‌های تابعی و به حرکت سیگنال‌ها در مسیر برگشت (از سمت راست به چپ)، سیگنال‌های خطا می‌گویند (شکل ۱).



شکل ۱. انتشار سیگنال در الگوریتم bp [۱۵]

۲-۴ درخت تصمیم

درخت تصمیم در یادگیری ماشین، مدل پیش‌بینی است که حقایق مشاهدات یک پدیده را به استنتاج‌هایی در مورد مقدار هدف آن پدیده ترسیم می‌کند. تکنیک یادگیری ماشین (به‌عنوان رایج‌ترین روش داده کاوی) برای استنتاج یک درخت تصمیم از داده‌ها، یادگیری درخت تصمیم نامیده می‌شود [۱۵، ۱۶]. هر گره داخلی، متناظر یک متغیر و هر کمان به یک فرزند، نمایان‌گر یک مقدار ممکن برای آن متغیر است. یک گره برگ، با داشتن مقادیر متغیرها که با مسیری از ریشه درخت تا آن گره برگ بازنمایی می‌شود، مقدار پیش‌بینی شده متغیر هدف را نشان می‌دهد. ساختار درخت تصمیم، به نوعی است که برگ‌ها، نشان‌دهنده دسته‌بندی و شاخه‌ها، ترکیبات فصلی صفاتی که منتج به این دسته‌بندی‌ها است، را بازنمایی می‌کنند. یادگیری یک درخت می‌تواند با تفکیک کردن یک مجموعه منبع، به زیرمجموعه‌هایی براساس یک تست مقدار صفت انجام شود. این فرآیند به شکل بازگشتی در هر زیرمجموعه حاصل از تفکیک تکرار می‌شود. عمل بازگشت زمانی کامل می‌شود که تفکیک بیشتر سودمند نبوده یا بتوان یک دسته‌بندی را به همه نمونه‌های موجود در زیرمجموعه به‌دست آمده، اعمال کرد [۱۶].

۳ مروری بر پیشینه پژوهش

میرعلوی و پورزمانی (۱۳۹۸)، در مطالعه‌ای به ارایه مدلی جهت پیش‌بینی قیمت سهام با استفاده از روش‌های فرااب‌تکاری و شبکه‌های عصبی پرداخته‌اند. در مدل پیشنهادی پژوهش، برای پیش‌بینی از یک سیستم دو سطحی از شبکه‌های عصبی پرسپترون چندلایه و چندین شاخص و برای انتخاب بهترین نمونه از الگوریتم بهینه‌سازی ملخ استفاده شده است. نتایج پژوهش نشان داده که مدل پیشنهادی توانسته با خطای پیش‌بینی پایین‌تری، نسبت به دیگر مدل‌ها عمل کند [۱۷].

همایون فر و همکاران (۱۳۹۷)، در مطالعه‌ای به بررسی توسعه الگوریتم‌های فرا اب‌تکاری شیرمورچه-ژنتیک و الگوریتم ترکیبی یادگیری افزایشی مبتنی بر جمعیت و تکامل تفاضلی^۱ جهت بهینه‌سازی سبد سهام در بورس اوراق بهادار تهران پرداخته‌اند. در این پژوهش برای بهینه‌سازی سبد سرمایه‌گذاری با هدف افزایش بازده و کاهش ریسک از الگوریتم ترکیبی شیرمورچه-ژنتیک و الگوریتم ترکیبی یادگیری افزایشی مبتنی بر جمعیت و تکامل تفاضلی استفاده شده است. نتایج پژوهش نشان داد، الگوریتم ژنتیک و الگوریتم ترکیبی یادگیری افزایشی مبتنی بر جمعیت و تکامل تفاضلی، کارایی مناسبی برای حل مساله بهینه‌سازی سبد سهام داشته و با استفاده از الگوریتم ژنتیک می‌توان سبد بهینه سهام را با دقت و کارایی بالا تشکیل داد [۷].

صادقی شریف، سیدجلال (۱۳۹۶) در پژوهشی به مطالعه ارزیابی پیش‌بینی پذیری قیمت سهام با استفاده از شبکه‌های عصبی فازی در بازار بورس تهران پرداخته است. این مطالعه نشان داده که شبکه‌های عصبی که مدل ساده‌ای از سیستم عصبی مرکزی انسان می‌باشند، روش‌های جدیدتری برای پیش‌بینی روند قیمت سهام ارایه

¹ Population-based incremental learning and differential evolution algorithm (PBILDE)

نموده‌اند. یک شبکه عصبی متشکل از نرون‌ها و گره یا نرون کوچک‌ترین واحد پردازش و توزیع اطلاعات، اساس عملکرد عصبی را تشکیل و ورودی و خروجی مخصوص خود را دارد [۱۸].

خنده خوش (۱۳۹۵) عوامل مؤثر در پیش‌بینی شاخص قیمت بورس تهران با بهبود الگوریتم بهینه‌سازی ملخ در انتخاب بهترین نمونه‌ها در مدل آموزش چندتایی شبکه عصبی را مطالعه نموده است. در این مدل یک سیستم دوسطحی از شبکه‌های عصبی پرسپترون چندلایه و چندین شاخص برای پیش‌بینی استفاده شده و برای آموزش بهتر شبکه عصبی و در نتیجه بهبود نتایج به دست آمده، از الگوریتم بهینه‌سازی ملخ برای انتخاب بهترین نمونه‌ها برای آموزش شبکه عصبی استفاده شده است و نتایج نشان داده که مدل پیشنهادی توانسته با خطای پیش‌بینی پایین‌تری نسبت به دیگر مدل‌ها عمل کند [۱۹].

فشاری و مظاهری فر (۱۳۹۵) در مطالعه‌ای به مقایسه الگوریتم‌های پیش‌بینی و بهینه‌سازی در بورس اوراق بهادار تهران پرداختند. در این مطالعه، از دو الگوریتم شبکه عصبی مصنوعی و شبکه فازی-عصبی به عنوان دو الگوریتم پیش‌بینی قیمت اوراق بهادار و از دو الگوریتم حرکت تجمعی ذرات و الگوریتم ژنتیک، روش کوادراتیک به منظور حل مساله بهینه‌سازی پرتفوی بدون محدودیت استفاده شده است. نتایج نشان داده که شبکه‌های عصبی نسبت به سیستم فازی عصبی، عملکرد بهتری را در پیش‌بینی بازده اوراق بهادار داشته و الگوریتم جهش ترکیبی قورباغه، عملکرد بهتری را در مقایسه با الگوریتم ژنتیک نسبت به الگوریتم کوادراتیک داشته است. بنابراین شبکه عصبی می‌تواند الگوریتم قابل اتکا برای سهامداران باشد [۲۰].

رستمی (۱۳۹۵) در مطالعه‌ای به بررسی کارایی الگوریتم یادگیری ماشین در پیش‌بینی شاخص بورس و اوراق بهادار تهران پرداختند. در مرحله اول متغیرهای مؤثر بر میزان تغییرات شاخص سهام بورس و در مرحله دوم با استفاده از دو مدل رافست و رگرسیون لجستیک، به کمک دو نرم افزار رزتا و وکا به پیش‌بینی شاخص بورس پرداخته‌اند. سپس مقادیر پیش‌بینی شده با این مدل‌ها را از طریق آزمون میانگین زوجی در نرم افزار SPSS با مقادیر واقعی مقایسه و نهایتاً با استفاده از آزمون آنوا برتری کارایی مدل‌ها را بررسی نموده‌اند. نتایج نشان داد که مدل‌های رافست و رگرسیون لجستیک (زیرمجموعه الگوریتم یادگیری ماشین)، توانایی پیش‌بینی شاخص سهام را دارند [۲۱].

تهرانی و همکاران (۱۳۹۴) در مطالعه‌ای به ارزیابی رویکردی جدید برای مدیریت فعال پرتفوی و انجام معاملات هوشمند سهام با تاکید بر نگرش انتخاب ویژگی پرداختند. در این تحقیق ابتدا با استفاده از روش مارکوویتز اوزان هریک از شش سهم پرتفوی ساخته شده در هریک از بازه زمانی چهار ساله را به دست آورده، سپس با استفاده از بیست و دو نماگر تکنیکی که به عنوان ورودی الگوریتم ژنتیک به عنوان یک روش انتخاب ویژگی در نظر گرفته شده است، استفاده کرده و به کمک دو روش پیش‌بینی مجاورت نزدیک‌ترین همسایه و شبکه عصبی اقدام به پیش‌بینی قیمت هر سهم نموده است. نتایج پژوهش نشان داده که پرتفوی حاصل از روش شبکه عصبی مبتنی بر الگوریتم ژنتیک دارای بالاترین بازده بوده که این خود نشان از برتری رویکرد مدیریت فعال پرتفوی نسبت به رویکرد مدیریت غیرفعال پرتفوی می‌باشد [۱].

یزدی و همکاران (۱۳۹۴) در مطالعه‌ای به مقایسه عملکرد چهار الگوریتم فراابتکاری نوین برای حل مسایل بهینه‌سازی ریاضی پرداختند. در این پژوهش با مقایسه الگوریتم بهینه‌سازی گرگ خاکستری با سه الگوریتم نوین بهینه‌سازی عملکرد، این الگوریتم‌ها از دو جهت دستیابی به نقطه بهینه سراسری و زمان دستیابی به این نقطه، مورد بررسی قرار گرفته است. به جهت صحیح بودن مقایسه، الگوریتم‌های انتخاب شده همگی از جمله الگوریتم‌هایی هستند که بر گرفته از رفتار غذایی موجودات زنده می‌باشند [۲۲].

یانگ سین و همکاران^۱ (۲۰۱۳) در مطالعه‌ای به بررسی الگوریتم PSO در برش سهام نامنظم پرداختند. مشکلات برش سهام (CSP) در بسیاری از صنایع تولیدی به وجود می‌آیند که در آن ورق‌های سهام بزرگ باید به قطعات کوچک‌تر برسد. در این تحقیق یک رویکرد تودرتو نامنظم برای مساله برش دو بعدی طراحی شده است. یک روش اکتشافی مبتنی بر الگوریتم بهینه‌سازی ازدحام ذرات (PSO) برای مساله برش دو بعدی در جهت یافتن راه حل بهینه ارائه داده است. علاوه بر این، رویکرد پیشنهادی یک روش تقریبی شبکه‌ای با استراتژی قرار دادن اکتشافی پایین سمت چپ را برای تخصیص موارد نامنظم ترکیب می‌کند. در این پژوهش، رویکرد پیشنهادی با استفاده از پانزده معیار بازنگری و ارزیابی شده است. عملکرد نشان‌دهنده اثربخشی و کارایی این رویکرد در حل مشکلات برش سهام نامنظم است [۱۱].

چن و همکاران (۲۰۱۱) در مقاله‌ای با عنوان الگوریتم ژنتیک رابطه‌ای به همراه جهش هدایت‌شده برای بهینه‌سازی پرتفوی مقیاس بزرگ، از الگوریتم ژنتیک رابطه‌ای به همراه یک عملگر جدید ارائه می‌نماید که جهش خوانده می‌شود. برای انتخاب موثرترین پرتفوی الگوریتم ژنتیک رابطه‌ای، ضرایب همبستگی بین برندهای سهام را به عنوان قدرت در نظر می‌گیرد که نشان‌دهنده رابطه بین گره‌ها در هر یک از الگوریتم ژنتیک رابطه‌ای می‌باشد. جهش هدایت‌شده پرتفوی جدید را مطابق با مقدار میانگین ضرایب همبستگی بین سهام تولید می‌کند که به معنی قابلیت بهره‌برداری از تکامل الگوریتم ژنتیک رابطه‌ای است. نتایج این تحقیق نشان داد که راهکار الگوریتم ژنتیک رابطه‌ای با جهش هدایت‌شده، موفق بوده است و پرتفوی به دست آمده در محدوده یا نزدیک به محدوده قابل قبول می‌گنجد [۲۳].

ایستون و پیتز (۲۰۱۶)، در مطالعه‌ای دریافتند که در برش‌های مقطعی از سهام در بورس نیویورک که از سال ۱۹۲۶ به طور مداوم معامله شده‌اند، نسبت قیمت به سود تقسیمی (P/D) یکی از قوی‌ترین پیش‌بینی‌کننده‌های ارزش فعلی تغییرات سودهای آتی است. بنابراین، نسبت (P/D) پیش‌بینی‌هایی از تغییرات بلند مدت در سودهای آتی ارائه می‌کند. این بدین معنی نیست که حساب‌های با اهمیتی در قیمت تک تک سهام وجود ندارد. به همان اندازه که حساب از دریا بیرون می‌آید، تغییر قابل پیش‌بینی در سود سهام شرکت‌ها اتفاق می‌افتد. تعداد زیادی از این تغییرات قابل پیش‌بینی در شرکت‌هایی است که سال‌های متمادی هیچ پرداخت سودی نداشتند، ولی سرمایه‌گذار به درستی در می‌یابد که سود، پرداخت خواهد شد [۲۴].

¹ Yanxin

خلاصه‌ای از کارهای انجام شده را می‌توان در جدول ۱ مشاهده نمود. با توجه به کارهای بررسی شده می‌توان دریافت هر یک از این روش‌ها دارای مزایا و معایبی می‌باشند، ولی در روشی که در این تحقیق قصد ارایه آن را داریم، سعی کردیم تا تمامی مزایای روش‌های مختلف را در کنار هم قرار دهیم و از معایب آن‌ها با استفاده از ترکیب کردن روش‌های قابل ترکیب، دوری کنیم.

جدول ۱. خلاصه‌ای از تحقیقات بررسی شده

نویسنده	موضوع	نتایج حاصل از تحقیق
خنده خوش (۱۳۹۵) [۱۹]	عوامل مؤثر در پیش‌بینی شاخص قیمت بورس تهران با بهبود الگوریتم بهینه‌سازی ملخ در انتخاب بهترین نمونه‌ها در مدل آموزش چندتایی شبکه عصبی	مدل پیشنهادی توانسته با خطای پیش‌بینی پایین‌تری نسبت به دیگر مدل‌ها عمل کند
همایون فرو همکاران (۱۳۹۷) [۷]	توسعه الگوریتم‌های فرا ابتکاری شیرمورچه- ژنتیک و الگوریتم ترکیبی یادگیری افزایشی مبتنی بر جمعیت و تکامل تفاضلی کارایی مناسبی برای حل مسأله بهینه‌سازی سبد سهام دارند. بعلاوه، با استفاده از الگوریتم فرا ابتکاری شیرمورچه- ژنتیک می‌توان با دقت و کارایی بالا سبد بهینه سهام تشکیل داد.	نتایج پژوهش حاکی از آن است که الگوریتم‌های فرا ابتکاری شیرمورچه- ژنتیک و الگوریتم ترکیبی یادگیری افزایشی مبتنی بر جمعیت و تکامل تفاضلی کارایی مناسبی برای حل مسأله بهینه‌سازی سبد سهام دارند. بعلاوه، با استفاده از الگوریتم فرا ابتکاری شیرمورچه- ژنتیک می‌توان با دقت و کارایی بالا سبد بهینه سهام تشکیل داد.
میرعلوی و پورزمانی (۱۳۹۸) [۱۷]	ارایه مدلی جهت پیش‌بینی قیمت سهام با استفاده از روش‌های فرا ابتکاری و شبکه‌های عصبی پرداخته‌اند.	نتایج حاصل از پژوهش نشان‌دهنده که مدل پیشنهادی توانسته با خطای پیش‌بینی پایین‌تری نسبت به دیگر مدل‌ها عمل کند.
فشاری و مظاهری فر (۱۳۹۵) [۲۰]	به مقایسه الگوریتم‌های پیش‌بینی و بهینه سازی در بورس اوراق بهادار تهران پرداختند.	نتایج حاصل از این تحقیق نشان می‌دهد که شبکه‌های عصبی توانسته اند عملکرد بهتری را در پیش‌بینی بازده اوراق بهادار نسبت به سیستم فازی عصبی نشان دهد.
رستمی (۱۳۹۵) [۲۱]	به بررسی کارایی الگوریتم یادگیری ماشین در پیش‌بینی شاخص بورس و اوراق بهادار تهران پرداختند.	نتایج حاصل این پژوهش نشان از آن است که مدل‌های رافت و رگرسیون لوجستیک که زیر مجموعه‌های الگوریتم‌های یادگیری ماشین هستند، توانایی پیش‌بینی شاخص سهام را دارند.
تهرانی و همکاران (۱۳۹۴) [۱]	به ارایه رویکردی جدید برای مدیریت فعال پرتفوی و انجام معاملات هوشمند سهام با تاکید بر نگرش انتخاب ویژگی پرداختند.	پرتفوی حاصل از روش شبکه عصبی مبتنی بر الگوریتم ژنتیک دارای بالاترین بازده در دوره چهار ساله را دارا بوده که این خود نشان از برتری رویکرد مدیریت فعال پرتفوی نسبت به رویکرد مدیریت غیرفعال پرتفوی می‌باشد.
یزدی و همکاران (۱۳۹۴) [۲۲]	به مقایسه عملکرد چهار الگوریتم فرا ابتکاری نوین برای حل مسایل بهینه‌سازی ریاضی پرداختند.	در این مقاله سعی شده است با مقایسه الگوریتم بهینه‌سازی گرگ خاکستری با سه الگوریتم نوین بهینه‌سازی عملکرد، این الگوریتم‌ها از دو جهت دستیابی به نقطه بهینه سراسری و همچنین زمان دستیابی به این نقطه، مورد بررسی قرار گیرد.
یانگ سین و همکاران (۲۰۱۳) [۱۱]	به بررسی الگوریتم PSO در برش سهام نامنظم پرداختند.	در این پژوهش، رویکرد پیشنهادی با استفاده از ۱۵ معیار بازنگری و ارزیابی می‌شود. عملکرد نشان‌دهنده اثربخشی و کارایی این رویکرد در حل مشکلات برش سهام نامنظم است.
چن و همکاران (۲۰۱۱) [۲۳]	در مقاله ای با عنوان الگوریتم ژنتیک رابطه ای به همراه جهش هدایت شده برای بهینه سازی پرتفوی مقیاس بزرگ که از مهم‌ترین	نتایج این تحقیق حاکی از آن است که راهکار الگوریتم ژنتیک رابطه ای با جهش هدایت شده، موفق بوده است و پرتفوی به‌دست آمده در محدوده یا نزدیک به محدوده قابل قبول می‌گنجد.

	تحقیقات اخیر در زمینه موضوع تحقیق می‌باشد.	
دریافتند که جاب‌های با اهمیتی در قیمت تک تک سهام وجود ندارد. این جاب‌ها بدین شکل می‌باشند که تغییرات قابل پیش‌بینی در شرکت در سال‌های متمادی دیده نمی‌شود ولی سرمایه‌گذار به درستی در می‌یابد که سود در نهایت پرداخت خواهد شد.	در این مقاله نسبت قیمت به سود تقسیمی (P/D) یکی از قوی‌ترین پیش‌بینی‌کننده‌های ارزش فعلی تغییرات سودهای آتی در نظر گرفته شده است.	ایستون و پیتر (۲۰۱۶) [۲۴]

۴ روش پیشنهادی

در پژوهش حاضر، هدف به کارگیری روشی است که به وسیله آن بتوان تا با درصد بالایی، عوامل موثر در بازار سهام را شناسایی و پیش‌بینی را انجام داد. در نهایت با استفاده از نتایج به‌دست آمده، راهکاری برای انتخاب بهترین سبد سهام ارائه نمود که بیشترین سود ممکن را کسب نماید. روش پیشنهادی پژوهش حاضر استفاده از الگوریتم ژنتیک بهبودیافته و شبکه‌های عصبی چند لایه است. در این پژوهش به کمک شبکه ژنتیک بهبودیافته و شبکه‌های عصبی تعداد ویژگی‌ها کاهش داده می‌شوند. با کاهش ویژگی، سربار محاسبات کاهش یافته و در نتیجه می‌توان سریع‌تر به نتیجه رسید و داده کاوی را انجام داد. بعد از مرحله کاهش ویژگی و آموزش سیستم پیشنهادی، از درخت تصمیم‌گیر پیشنهادی که در راستای ساخت درخت بهینه شده است، استفاده می‌شود تا بتوان با استفاده از روابط منطقی، تصمیمات خروجی بیان شود. این تصمیمات با توجه به قواعد ریاضی، استخراج می‌شود و دارای دقت بالایی می‌باشند و با استفاده از این سیستم پیشنهادی می‌توان حتی آن را هوشمند نیز نمود تا بتواند در شرایط خاص با توجه به تصمیمات استخراج شده خودکار، توصیه‌های خودکاری را نیز بسته به شرایط انجام داد. مراحل مختلف روش پیشنهادی به ترتیب عبارتند از: ۱- پیش‌پردازش و انتقال داده‌ها، ۲- کاهش ویژگی و آموزش با استفاده از الگوریتم شبکه ژنتیک بهبودیافته و شبکه‌های عصبی، ۳- ساخت درخت تصمیم‌گیری، ۴- تبدیل درخت تصمیم‌گیری، ۵- استخراج تصمیمات.

۴-۱ پیش‌پردازش و انتقال داده‌ها

در ابتدا مجموعه داده‌ها، جمع‌آوری شده و سپس آماده‌سازی و پیش‌پردازش داده‌ها انجام شده است. در آماده‌سازی و پیش‌پردازش داده‌ها از روش‌های مختلفی استفاده شده است. اول این که برخی ویژگی‌ها که دارای مقادیر منحصر به فرد بوده و از یک سو نمی‌توانند دانش مفیدی را در مجموعه داده‌ها ایجاد کنند و از سوی دیگر در صورت عدم حذف به دلیل پردازش، سربار محاسباتی را بالا می‌برند، از مجموعه داده‌ها حذف شده‌اند. از جمله ویژگی‌های داده‌ها که فاقد ارزش پردازش بوده و از مجموعه داده‌ها حذف شده‌اند می‌توان به ویژگی نام سهم و تاریخ، تراکنش‌های دارای مقادیر مفقود فراوان، اشاره نمود. از طرفی مقادیر برخی از ویژگی‌های مجموعه داده‌ها که دارای مقادیر نوین و مفقود بوده اصلاح شده‌اند. در مرحله بعدی از ابزار کشف آنومالی استفاده شده است. در این فرآیند داده‌هایی که در محدوده‌های خارج از محدوده مجاز قرار دارند، شناسایی و سپس حذف شده‌اند. برای این که بتوان روی داده‌ها به عنوان ورودی کار کرد، پیش از انتخاب و استخراج

ویژگی‌ها، برخی عملیات پیش پردازش بر روی داده‌ها انجام شده است. در مرحله انتقال داده‌ها به سیستم، داده‌هایی که خارج از دامنه بوده، به عنوان داده‌های مشکل دار، حذف شده‌اند.

۴-۲ انتخاب ویژگی‌های موثر با استفاده از الگوریتم ژنتیک و شبکه‌های عصبی

الگوریتم ژنتیک، به عنوان یکی از ابزارهای داده کاوی برای طبقه‌بندی داده‌ها و با قابلیت قابل توجه در استنتاج معانی از داده‌های پیچیده، برای استخراج الگوها و شناسایی روش‌هایی که آگاهی از آن‌ها برای انسان و دیگر تکنیک‌های کامپیوتری بسیار پیچیده و دشوار است، به کار گرفته شده است. در این قسمت سیستم مورد آزمایش با استفاده از الگوریتم ژنتیک مورد آموزش قرار گرفته و از طرفی دیگر شبکه‌های عصبی پرسپترون چند لایه باعث آموزش بسیار سریع روش پیشنهادی شده است. از آنجا که یکی از قابلیت‌های بسیار قوی شبکه‌های عصبی، امکان یادگیری با داده‌های کم و دقت نسبتاً خوب می باشد، در این پژوهش از ترکیب الگوریتم‌های ژنتیک و شبکه‌های عصبی پرسپترون چند لایه استفاده شده است تا به کمک آن دستیابی به راهکاری که بتوان به وسیله آن با داده‌های مختلفی کار کرد، امکان پذیر باشد. برای انجام پیش‌بینی، وزن‌هایی که در شبکه عصبی قرار گرفته با استفاده از الگوریتم ژنتیک، مقداردهی شده‌اند تا اینکه دقت روش پیشنهادی افزایش یابد.

تعداد نورون بالا در شبکه، موجب افزایش پیچیدگی شبکه شده و متعاقب آن شبکه دچار بیش برآزش شده و به تبع آن قابلیت پیش‌بینی شبکه تحت تاثیر قرار می‌گیرد. برای حل این موضوع بابت هر کدام از الگوریتم‌های آموزشی، تعداد نورون‌ها از یک تا ۱۰ افزایش داده شدند. داده‌ها در دامنه ۱- تا ۱ نرمالیزه شده و از تابع انتقالی تانژانتی - سیگموئیدی^۱ استفاده شده است. از دو شاخص ارزیابی عملکرد شبکه، خطای میانگین مربعات (MSE) و ضریب تبیین (R2) بهره گرفته شده است. الگوریتم ژنتیک به عنوان رایج‌ترین الگوریتم بهینه‌سازی که بر اساس دگرگونی بیولوژیکی به وجود آمده است، جمعیت در آن الگوریتم شامل پاسخ‌های ممکن در فرم آرایه‌ای از کروموزوم‌ها می‌باشد. بنابراین هر یک از جمعیت‌ها به صورت تصادفی به عنوان وزن شبکه توسط الگوریتم ژنتیک بهینه‌سازی شده‌اند. تابع MSE به عنوان تابع هزینه معرفی شده و کروموزوم‌های جمعیت سپس برای رسیدن به کمترین تابع هزینه مرتب شده‌اند. تعداد مشخصی از اعضای بهتر بر اساس کمترین هزینه به نسل بعدی منتقل شده‌اند. در این مرحله سه اپراتور الگوریتم ژنتیک (انتخاب، تقاطع و دگرگونی) برای تولید جمعیت نسل بعدی فعال شده‌اند. این سیکل تا رسیدن به جواب مطلوب و نیل به وزن‌های مطلوب شبکه ادامه یافته است.

۴-۳ ساخت درخت تصمیم‌گیری

درخت تصمیم‌گیری، درختی است که هر شاخه از آن به عنوان یک انتخاب می‌باشد. بدین معنی برای رفتن از گره ریشه به گره پایین‌تر می‌توان از شاخه‌هایی که به آن گره متصل هستند، یکی انتخاب شود. در انتها هر یک از گره‌های انتهایی یا اصطلاحاً گره برگ تصمیمی را بازگو می‌کند. هر کدام از شاخه‌ها تا رسیدن به برگ دارای سناریویی می‌باشد که موجب اتخاذ یک تصمیم می‌شود. در این پژوهش از مدل پیشنهادی مبتنی بر درخت

¹ tansig

تصمیم، ID3 بهبود یافته استفاده شده است که این بهبود موجب سرعت عمل بالای آن شده است. درخت ID3 یک درخت تصمیم گیری است که دارای یادگیری نیز می باشد و اولین بار توسط راس کوینلن^۱ مطرح شد. ایده الگوریتم ID3، ساخت درخت تصمیم گیری بالا به پایین می باشد که انتخاب گره در آن به وسیله جستجوی حریصانه از میان مجموعه ای از صفت ها می باشد. در این پژوهش برای یافتن مفیدترین صفت از میان صفاتی که در کلاسه بندی مفیدتر باشد، از الگوی به خصوصی استفاده شده است. برای اینکه بتوان کلاسه بندی مفیدی را برای مجموعه یادگیری انجام داد، تعداد سوالات را کاهش داده یعنی عمق درخت تصمیم گیری کاهش داده شده است. از این رو در این قسمت نیاز به تابعی است که بتوان متعادل ترین تقسیم را انجام داد تا عمق درخت بسیار کاهش یابد و گره ها به صورت متعادل در درخت تقسیم شوند.

توابع زیادی چون آنتروپی، gini index و classification error برای سنجش میزان همگن پذیری وجود دارند. از این میان در پژوهش حاضر از آنتروپی^۲ استفاده شده است.

$$Entropy = \sum_j -p_j \log_2 p_j \quad (1)$$

آنتروپی یک جدول صفر است، زیرا احتمال آن مقداری برابر یک است (تنها دارای یک کلاس باشد). آنتروپی زمانی به بیشترین مقدار خود می رسد که تمامی کلاس های موجود در جدول دارای احتمالی برابر باشند. آنتروپی را می توان به نوعی معیاری برای سنجش بی نظمی در نظر گرفت. هر چه مجموعه منظم تر و دارای گوناگونی کمتری باشد، آنگاه آنتروپی آن کم تر است و به نوعی بی نظمی آن نیز کم تر خواهد شد و برعکس. البته در اینجا چون در مرحله قبل یک کلاسه بندی ابتدایی را انجام داده ایم، تقریباً بی نظمی نیز پایین می باشد و این خود باعث سرعت عمل بالاتر روش پیشنهادی ما شده است. زیرا این قضیه باعث شده تا عمق درخت تصمیم گیری کم شود و هر چه عمق این درخت کم تر شود، سرعت تصمیم گیری نیز بیشتر می شود.

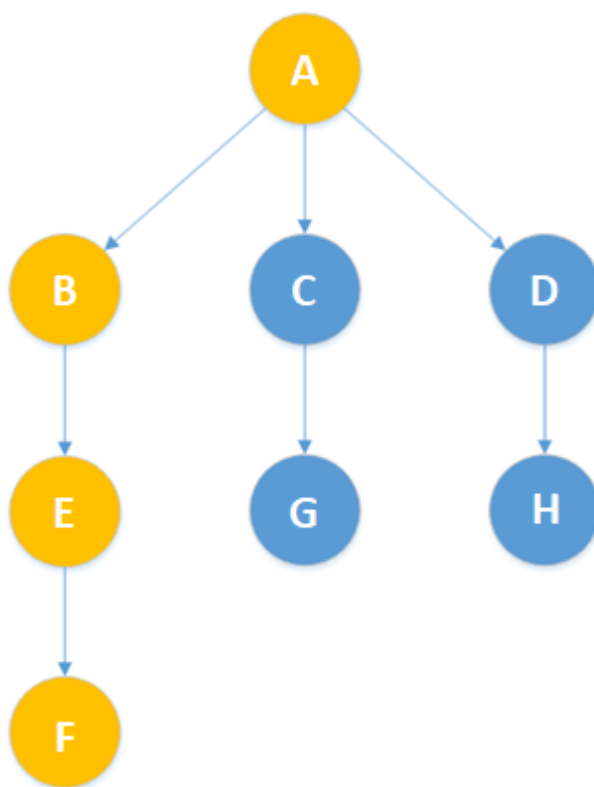
در این قسمت ما از آنتروپی استفاده کرده ایم تا مقدار بی نظمی را برای هر صفت از جدول ۲ به دست آوریم. برای این که بتوانیم صفتی را در درخت تصمیم گیری انتخاب کنیم که در رتبه بالاتری از بقیه صفت ها باشد و به نوعی دارای اهمیت بالاتری از بقیه صفت ها باشد، از فرمول (۱) استفاده می کنیم. با توجه به این فرمول ما آنتروپی همه صفات را در مجموعه S محاسبه کرده و مقدار صفت مجموعه A را از آن کاسته ایم. مجموعه A، مجموعه همه صفات انتخاب شده از پدر تا به اینجا در یک مسیر خاص می باشد.

$$G(S, A) = Entropy(S) - \sum_{v \in \text{values}(A)} Entropy(v) \quad (2)$$

برای درک بهتر فرمول (۲) می توان شکل ۲ را مشاهده نمود.

¹ Ross Quinlan

² Entropy



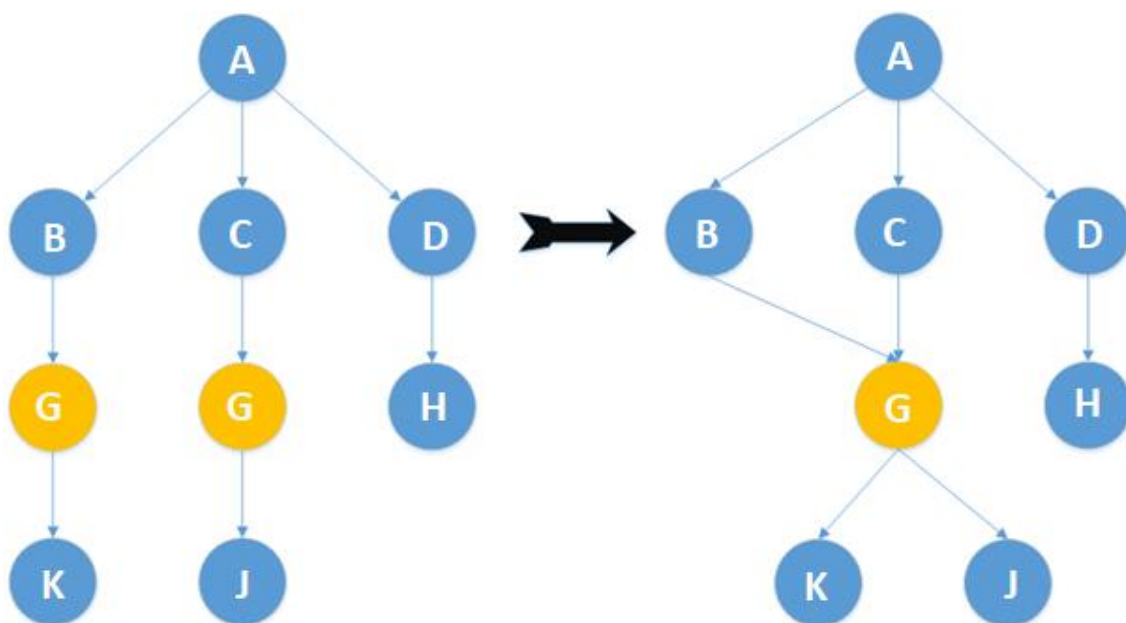
شکل ۲. مثالی برای انتخاب صفت‌ها

همان‌طور که در شکل ۲ قابل مشاهده است، ما آنتروپی تمامی صفات را از آنتروپی صفات انتخاب شده تا به اینجای مسیر کاسته‌ایم (یعنی $G(S,F)=E(S)-(E(A)+E(B)+E(E)+E(F))$ را به دست آورده‌ایم). در مجموعه صفات استفاده شده تا به اینجا کار به همراه صفاتی که قرار است از این مرحله مورد محاسبه قرار گیرند در این مجموعه قرار می‌گیرند. در این مجموعه برای هر یک از اعضا وزن محاسبه می‌شود که برای این کار از فرمول (۲) استفاده می‌شود. از این بین صفت و یا عضوی از مجموعه که دارای بیشترین Gain است انتخاب می‌گردد. در حالتی که دو صفت یا دو عضو دارای Gain برابری باشند که احتمال این پیشامد نیز کم نیست، به گره دو یا هر تعداد صفت که دارای بیشترین مقدار Gain هستند و باهم برابر نیز می‌باشند، به عنوان یک فرزند گره حاضر در نظر گرفته می‌شوند. یعنی اگر برای مثال در گره‌ای دو صفت دارای G برابر بودند، آنگاه این دو به گره مربوطه دو فرزند افزوده شده و هر کدام از این صفات‌ها به عنوان یک فرزند این گره در نظر گرفته شده و بعد از این کار، روند الگوریتم برای هر یک از این گره‌ها ادامه یافته است. برای مثال در شکل ۲ می‌توان مشاهده نمود که مقدار G برای صفات‌های B، C و D برابر است. از این رو همه این صفات‌ها در یک سطح قرار گرفته‌اند. این کار باعث شد تا صفت‌هایی که دارای آنتروپی بیشتری هستند، را بیاییم. چراکه این صفات‌ها تاثیر بیشتری را در تصمیم نهایی ما گذاشته‌اند. این روند جلو رفتن در درخت تصمیم‌گیری تا جایی ادامه یافت که در هر مسیر دیگر صفتی باقی نمانده باشد. در این حالت درخت تصمیم‌گیری کاملاً ساخته شده و به پایان رسیده است.

۴-۴ تبدیل درخت تصمیم گیری

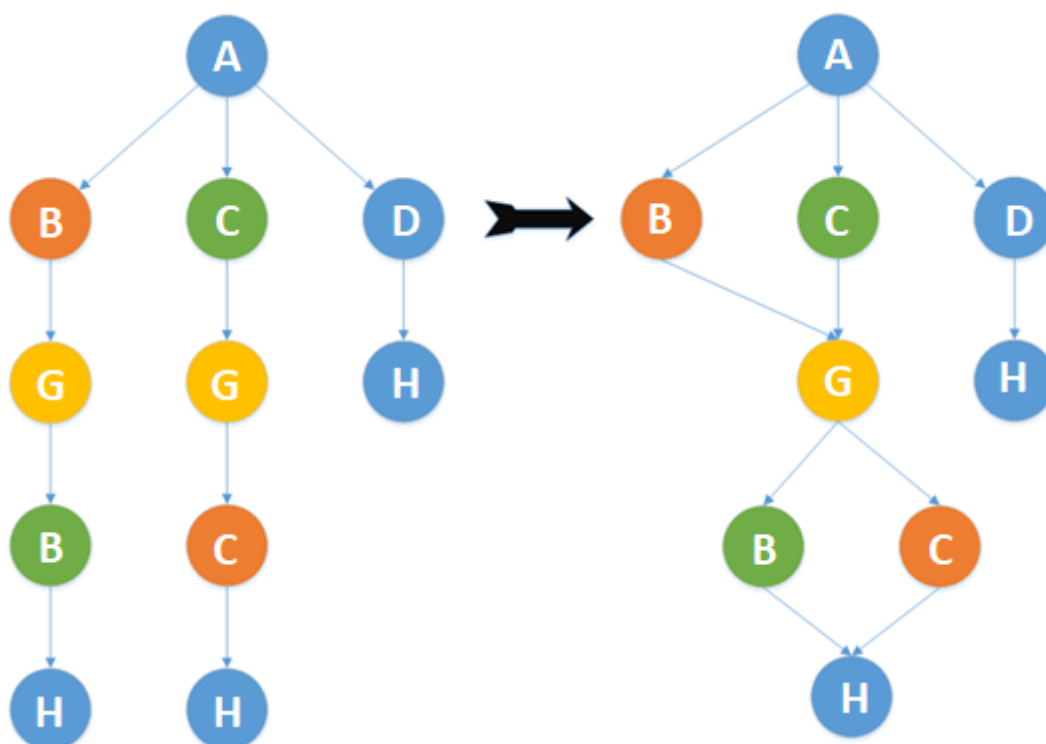
برای اینکه بتوان شرط‌هایی را از درخت تصمیم استخراج نمود (به صورت If...then....)، و بتوان از درخت تصمیم و قوانین انجمنی نیز کمک گرفت، ما درخت را تبدیل نمودیم. در موارد بسیاری این درخت تصمیم از حالت درخت، خارج شده و تبدیل به گراف شده است. اگر چه این درخت از نظر ظاهری تبدیل به گراف شده است، ولیکن برای ما همچنان به صورت درخت در نظر گرفته شده است. بنابراین می‌توان گفت در تبدیل درخت تصمیم، چیزی مابین گراف و درخت به دست آمده است.

در این قسمت ما در هر سطح، گره‌های همنام را با هم ادغام کرده و فرزندان آن‌ها به این گره ادغام شده، افزوده شده‌اند. می‌توان در شکل ۳ این کار را مشاهده نمود تا بهتر بتوان آن را درک نمود.



شکل ۳. مثالی از تبدیل درخت تصمیم گیری

همان‌طور که در شکل ۳ می‌توان مشاهده نمود، گره‌های B و C دارای فرزند مشترک G هستند. این دو گره در یک سطح نیز قرار دارند. در این حالت دو گره، تبدیل به یک گره شده و فرزندان آن‌ها نیز به گره جدید افزوده شده‌اند. ممکن است شرایط مانند شکل ۴ پیش آید که در آن دو گره G با هم ادغام شوند. ولی حالتی است که در آن گره G اول دارای فرزند C و گره G دوم دارای فرزند B باشد و این گره‌ها هر کدام برای گره G به منزله فرزند به حساب می‌آیند. در این حالت نیز مشکلی پیش نمی‌آید و تنها می‌توان گفت در هنگام نوشتن شرط می‌توان در نظر گرفت که با داشتن شرط B و C، همراه با and منطقی آن با G آن مسیر پیموده می‌شود. یعنی $A \text{ and } (B \text{ or } C) \text{ and } G$.



شکل ۴. مثالی از حالت خاص تبدیل

در اینجا می‌رسیم به این قسمت که هر کدام از این صفت‌ها در هر یک از داده‌ها می‌توانند دارای مقدار متفاوتی باشند، برای مثال در شکل ۵، G می‌تواند دارای دو مقدار true و یا false باشد، برای این قسمت ما داده‌ای را انتخاب می‌کنیم که فراوانی آن در صفت مربوطه در کل داده‌ها بیشتر است. یعنی به عنوان مثال اگر در بین صفت داده‌های از نوع G، فراوانی true بیشتر باشد یا false، آنگاه آن را بر می‌گزینیم. در صورتی که فراوانی دو مقدار برابر باشد، آنگاه در تصمیم، or را برای این دو مقدار قید می‌کنیم. برای مثال $G=true$ or $G=false$. برای استخراج تصمیمات از گراف تصمیم‌گیری که در اینجا ساخته شده، از گره ریشه شروع کرده و به سمت هر یک از برگ‌ها می‌رویم. در این صورت یک تصمیم شکل می‌گیرد.

دیتاست‌های مربوط به بازار بورس جهانی را دارد که در پژوهش حاضر از مجموعه‌ای از آن با نام Tata Global Beverages استفاده شده است. داده‌های استفاده شده مربوط به سال ۲۰۱۸ می‌باشند. این مجموعه شامل مجموعه اطلاعاتی می‌باشد که در جدول ۲ مشخص شده است.

جدول ۲. مشخصات مجموعه داده Tata Global Beverages

Date	Open	High	Low	Last	Close	Total trade Quantity	Turnover (Lacs)
۲۰۱۸-۱۰-۰۸	۲۰۸,۰۰	۲۲,۲۵	۰۶,۸۵	۲۱۶,۰۰	۲۱۵,۱۵	۴۶۴۲۱۴۶,۰	۱۰۰۶۲,۸۳
۲۰۱۸-۱۰-۰۵	۲۱۷,۰۰	۱۸,۶۰	۰۵,۹۰	۲۱۰,۲۰	۲۰۹,۲۰	۳۵۱۹۵۱۵,۰	۷۴۰۷,۰۶
۲۰۱۸-۱۰-۰۴	۲۲۳,۵۰	۲۷,۸۰	۱۶,۱۵	۲۱۷,۲۰	۲۱۸,۲۰	۱۷۲۸۷۸۶,۰	۳۸۱۵,۷۹
۲۰۱۸-۱۰-۰۳	۲۳۰,۰۰	۳۷,۵۰	۲۵,۷۵	۲۲۶,۴۰	۲۲۷,۶۰	۱۷۰۸۵۹۰,۰	۳۹۶۰,۲۷
۲۰۱۸-۱۰-۰۱	۲۳۴,۵۵	۳۴,۶۰	۲۱,۰۵	۲۳۰,۳۰	۲۳۰,۹۰	۱۵۳۴۷۴۹,۰	۳۴۸۶,۰۵

ستونهای open و close نشان‌دهنده قیمت پایه و نهایی است که سهام آن در یک روز خاص معامله می‌شود. low high و last نشان‌دهنده حداکثر، حداقل و آخرین قیمت سهم در آن روز است. Total trade quantity، تعداد سهام خریداری شده یا فروخته شده در روز و turnover، گردش مالی شرکت خاص در یک تاریخ معین است.

در این دیتاست، بعضی از مقادیر موجود نیستند، زیرا روزهایی که تعطیل رسمی است، مثل آخر هفته و یا تعطیلات دیگر، بازار بورس نیز تعطیل می‌باشد.

محاسبه سود یا ضرر معمولاً به وسیله قیمت close یک سهم برای هرروز مشخص می‌شود، بنابراین در این پژوهش ستون close به عنوان ستون هدف بوده و دارای اهمیت بالایی است.

در شکل ۶ می‌توان نمودار تغییرات قیمت را برای این دیتاست مشاهده نمود. در این پژوهش داده‌ها به دو قسمت، داده‌های تست و آموزش تقسیم شده‌اند که برای این کار از cross k fold با مقدار $k=10$ استفاده شده است. در شکل ۶ تغییرات قیمت در تاریخ نشان داده شده است. این نمودار مقادیر واقعی می‌باشد و در این پژوهش هدف ارایه راهکاری می‌باشد که بتوان نزدیک‌ترین تغییرات با مقدار واقعی را پیش‌بینی نمود.

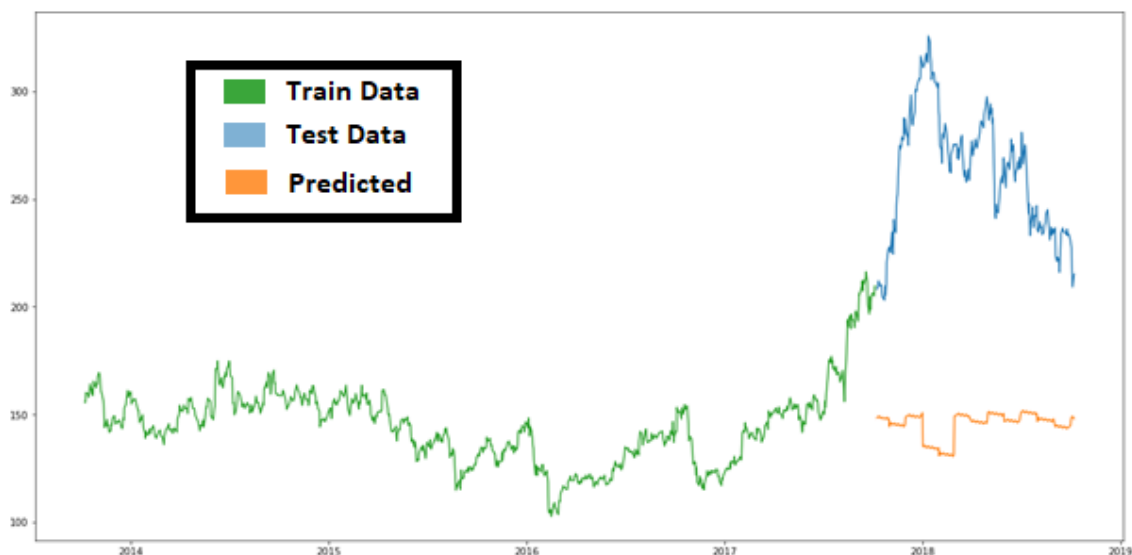


شکل ۶. نمودار تغییرات قیمت برای دیتاست مورد استفاده

برای اینکه بتوان مقدار پیش‌بینی شده را مورد ارزیابی قرار داد و میزان خطا را محاسبه نمود، مقدار تغییرات مشخص شده و مقدار پیش‌بینی شده با مقدار واقعی مورد مقایسه قرار داده شده است. لذا برای درک بهتر قدرت روش پیشنهادی، ذیلاً این روش با روش‌های مختلف دیگر مورد مقایسه قرار گرفته است.

۵-۱ رگرسیون خطی

در مقاله [۲۶] از رگرسیون خطی برای پیش‌بینی بازار بورس استفاده شده است که در این پژوهش نیز پیاده‌سازی شد تا بتوان دقت آن را با روش پیشنهادی با یک دیتاست مشابه بررسی نمود. نتیجه به دست آمده از این روش را می‌توان در شکل ۷ برای دیتاست استفاده شده مشاهده نمود. در این شکل رنگ سبز نشان دهنده مجموعه داده‌های آموزش می‌باشد و رنگ آبی مقادیر واقعی داده‌های تست می‌باشند. رنگ نارنجی، مقادیر پیش‌بینی شده توسط الگوریتم می‌باشند.

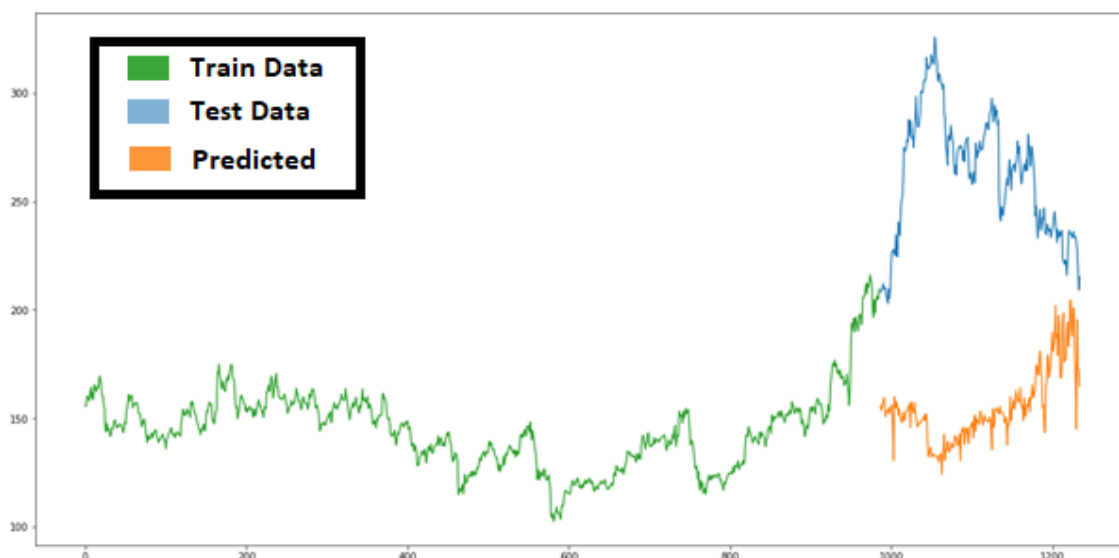


شکل ۷. نتیجه پیش‌بینی با استفاده از روش مقاله [۲۶]

در این قسمت نتایج رگرسیون خطی را می‌توان مشاهده نمود که با مقدار واقعی فاصله بسیاری دارد و اصلاً به نتایج واقعی نزدیک نمی‌باشد و این نشان می‌دهد که با توجه به ماهیت رگرسیون خطی، این روش قاعدتاً نمی‌تواند مقادیر بسیار نزدیکی را نسبت به مقدار واقعی پیش‌بینی نماید. یعنی داده‌هایی که دارای فراز و نشیب زیادی هستند، رگرسیون خطی جواب‌گو نمی‌باشد، ولی اگر یک سیر زمانی صعودی و یا نزولی به عنوان داده‌های آموزش داده می‌شد آنگاه قطعاً رگرسیون با توجه به سیر صعودی و یا نزولی آموزش دیده می‌تواند روند بسیار نزدیک‌تری پیش‌بینی نماید و در نتیجه در تست‌ها جواب‌های بهتری را به عمل آورد. در این آزمایش سعی شد تا داده‌هایی که برای آموزش داده می‌شود دارای فراز و نشیب‌های زیادی باشد که بتوان روش‌ها را در مقابل روش پیشنهادی بهتر مورد مقایسه قرار داد.

۲-۵ جستجوی نزدیک‌ترین همسایگی

در این پژوهش برای پیش‌بینی بازار بورس از روش پیشنهادی مقاله [۱۳] استفاده و پیاده‌سازی شده است که می‌توان در شکل ۸ نتیجه اجرای این روش روی دیتاست مشاهده نمود. در این شکل رنگ سبز نشان‌دهنده مجموعه داده‌های آموزش می‌باشد و رنگ آبی مقادیر واقعی داده‌های تست می‌باشند. رنگ نارنجی، مقادیر پیش‌بینی شده توسط الگوریتم می‌باشند.

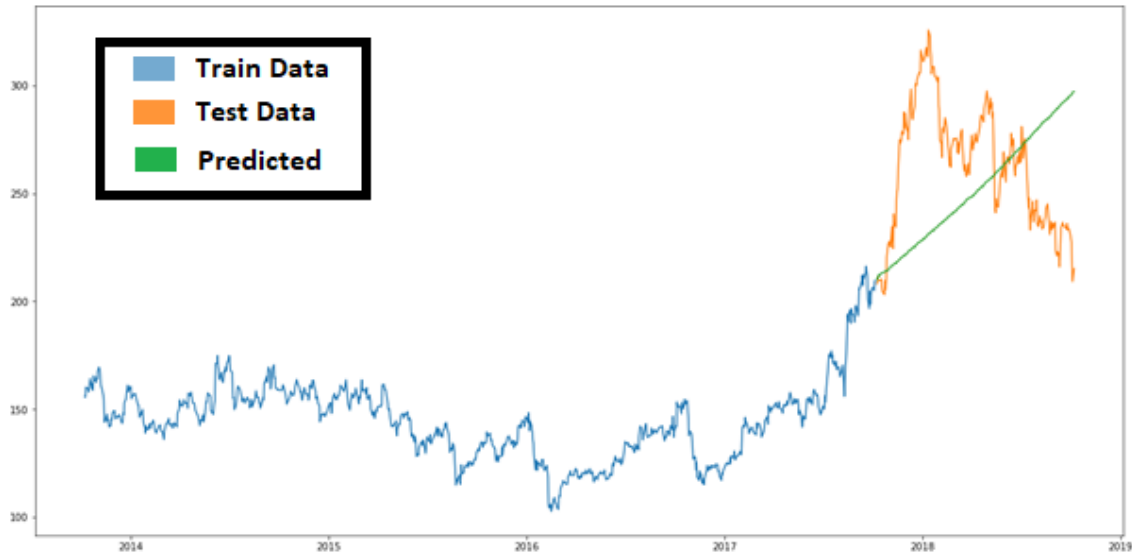


شکل ۸. نتیجه پیش‌بینی با استفاده از روش مقاله [۱۳]

در اینجا می‌توان مشاهده نمود که نتایج به دست آمده از رگرسیون خطی یعنی مقاله [۲۷] بسیار بهتر می‌باشد ولی همچنان با مقدار واقعی فاصله بسیاری را دارد.

۳-۵ میانگین متحرک خودهمبسته یکپارچه

در آمار و اقتصادسنجی و به ویژه در آنالیز سری‌های زمانی، میانگین متحرک خودهمبسته یکپارچه، مدلی گسترده‌تر از میانگین متحرک خودهمبسته است. این مدل‌ها در سری‌های زمانی برای فهم بهتر مدل یا پیش‌بینی آینده و درموردی که داده‌ها غیرایستا باشند، به کار می‌روند. در این حالت با یک بار دیفرانسیل‌گیری، جز یکپارچه بودن این داده‌ها از بین می‌رود و امکان برآورد یک ARMA در داده‌های جدید به وجود می‌آید. در مقاله [۱۷]، از روش ARIMA برای پیش‌بینی بازار بورس استفاده شده است به همین دلیل در این مقاله این روش نیز برای مقایسه پیاده‌سازی شده است که می‌توان در شکل ۹ نتیجه خروجی این روش را مشاهده نمود. در این شکل رنگ آبی نشان‌دهنده مجموعه داده‌های آموزش می‌باشد و رنگ نارنجی مقادیر واقعی داده‌های تست می‌باشند. رنگ سبز، مقادیر پیش‌بینی شده توسط الگوریتم می‌باشند.



شکل ۹. نتیجه پیش‌بینی با استفاده از روش مقاله [۲۷]

مشاهده می‌شود که در اینجا پیش‌بینی‌ها نسبت به روش‌های قبلی بهتر بوده، ولیکن نمی‌تواند نتایج دقیقی را مشخص کند. اما خروجی آن به‌طور میانگین با میانگین مقدار واقعی بسیار نزدیک است. می‌توان دریافت که راهکار مقاله [۲۶] بهتر است و قادر است تا پیش‌بینی‌های بهتری را انجام دهد. اما قابل توجه است که در پیش‌بینی‌ها، فراز و نشیب‌ها به درستی ترسیم نشده و این یک مشکل اساسی می‌باشد.

۵-۴ الگوریتم پیش‌بینی سری زمانی^۱

الگوریتم پیش‌بینی سری زمانی، توسط فیسبوک در جهت پیش‌بینی سیر زمانی طراحی و پیاده‌سازی شده است [۲۷]. الگوریتم پیش‌بینی سری زمانی یک کتابخانه پیش‌بینی سری زمانی است که نیاز به هیچ پیش‌پردازش داده‌ای ندارد. پیچیدگی محاسباتی آن پایین و سربار محاسباتی پایینی دارد.

در اینجا الگوریتم پیش‌بینی سری زمانی به عنوان یک الگوریتم معمول در دنیا، نیز پیاده‌سازی شده است و نتیجه خروجی آن در شکل ۱۰ نشان داده شده است. در این شکل رنگ سبز نشان‌دهنده مجموعه داده‌های آموزش می‌باشد و رنگ آبی مقادیر واقعی داده‌های تست می‌باشند. رنگ نارنجی، مقادیر پیش‌بینی شده توسط الگوریتم می‌باشند.

¹ Time series prophet algorithm

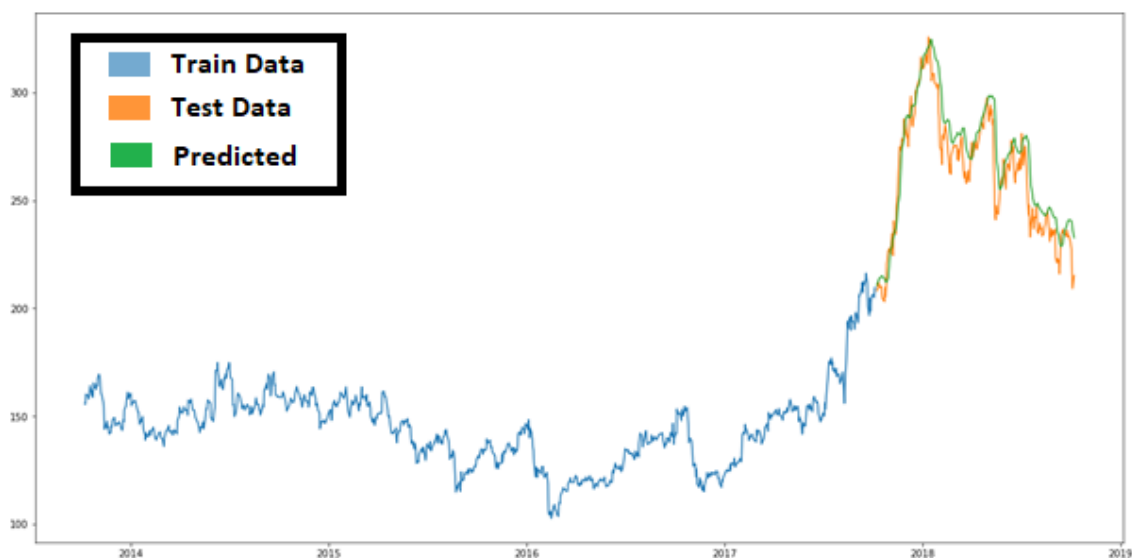


شکل ۱۰. نتیجه پیش‌بینی با استفاده از روش الگوریتم پیش‌بینی سری زمانی

مشاهده می‌شود که نتایج به دست آمده حاکی از آن است که این روش نسبت به روش‌های قبلی که تاکنون مورد بررسی قرار گرفته، دارای دقت بالاتری بوده و قادر است تا پیش‌بینی‌های نزدیک‌تری را ارائه نماید. همچنین می‌توان مشاهده نمود که این روش قادر است تا فراز و نشیب‌های تغییرات قیمت را بسیار بهتر از راهکارهای گذشته نشان دهد، اما پیش‌بینی‌های انجام شده از مقاله [۲۷] بهتر نبوده اما می‌توان مشاهده نمود که توانسته است فراز نشیب‌های مقادیر را برای داده‌های تست بسیار بهتر پیش‌بینی نماید.

۵-۵ خروجی روش پیشنهادی

نتیجه خروجی روش پیشنهادی که روی دیتاست مشخص شده، در شکل ۱۱ قابل مشاهده است. در این شکل رنگ آبی نشان‌دهنده مجموعه داده‌های آموزش می‌باشد و رنگ نارنجی مقادیر واقعی داده‌های تست می‌باشند. رنگ سبز، مقادیر پیش‌بینی شده توسط الگوریتم پیشنهادی می‌باشند.



شکل ۱۱. نتیجه پیش‌بینی با استفاده از روش پیشنهادی

همان‌طور که در نمودار بالا مشخص است، روش پیشنهادی از دیگر روش‌های مورد مطالعه دارای پیش‌بینی بهتری بوده و می‌توان به این نتیجه رسید که روش پیشنهادی قادر است تا با دقت بالاتری نسبت به روش‌های مورد بررسی، پیش‌بینی‌ها را انجام دهد. در جدول ۳ خطای پیش‌بینی برای روش‌های مختلف، قابل مشاهده است.

جدول ۳. خطای پیش‌بینی محاسبه شده برای روش‌های مختلف

روش	خطای پیش‌بینی (RMSE)
رگرسیون خطی	۱۲۱,۱۶۲۹۱۵۹۶۵۲۳۱۵۶
جستجوی نزدیک‌ترین همسایگی	۱۱۵,۱۷۰,۸۶۵۵۰,۰۲۶۷۲۱
میانگین متحرک خودهمبسته یکپارچه	۴۴,۹۵۴۵۸۴۹۹۳۲۴۶۹۵۴
الگوریتم پیش‌بینی سری زمانی	۵۷,۴۹۴۴۶۱۹۳۰۵۷۵۱۴۹
روش پیشنهادی	۱۱,۷۷۲۲۵۹۶۰۸۹۶۲۶۴۲

با توجه به اینکه این روش پژوهش قادر است تا بهترین پیش‌بینی را برای آینده انجام دهد، از این‌رو در این پژوهش می‌توان به راحتی با اعمال الگوریتم پیش‌بینی روی تمامی سهام موجود و پیش‌بینی تغییرات آینده بهترین سهام را شناسایی و بدین شکل بیش‌ترین مقدار سوددهی را مشخص نموده و کاربران می‌توانند به سود قطعی دست یابند. بنابراین در این قسمت تنها سهامی انتخاب می‌شوند که میانگین تغییرات قیمت آن‌ها از آستانه‌ای که توسط کاربر مشخص می‌شود، بالاتر باشد و در این حالت سبد سهام صورت می‌گیرد و حتی می‌توان تعداد سهام مورد انتخاب را نیز مشخص نمود که بهترین سبد سهام انتخاب و در نتیجه به بیشترین سود دست یافت.

۶ نتیجه‌گیری

در این تحقیق، یک روش مبتنی بر شبکه‌های عصبی، برنامه‌نویسی شبکه ژنتیک و درخت تصمیم برای پیش‌بینی قیمت و روند تغییر قیمت سهام جهت انتخاب سبد بهینه سهام ارائه شده است. از طرفی تحقیقات اخیر نشان داده است که طراحی و ارائه مدل‌های قاعده محور برای زمان خرید و فروش امکان‌پذیر بوده و می‌توان با تولید قواعد معاملاتی، سیستم‌های توانمندی را برای پشتیبانی از تصمیم‌گیری‌های سرمایه‌گذاران، توسعه داد. در گذشته راهکارهای مختلفی در راستای انتخاب سبد سهام ارائه شدند که هر یک دارای مزایا و معایبی هستند ولی از مهم‌ترین پارامترهای تاثیرگذار دقت است که راهکارهای گذشته دقت کافی را در پیش‌بینی روند بازار نداشتند و از این‌رو قادر نبودند تا با دقت بالایی سبد بهینه سهام را انتخاب نمایند. در تحقیق حاضر سعی شده است تا راهکاری ارائه شود که علاوه بر دقت بالا دارای عملکرد بسیار بالایی نیز باشد. روش ارائه شده در تحقیق حاضر دارای سربار حافظه‌ای و محاسباتی بسیار پایینی است. این روش قادر است تا برای هر سهم پیش‌بینی نماید که روند قیمتی سهم در آینده به چه صورت خواهد بود و از این جهت می‌توان به سودمندی و یا عدم سودمندی یک سهم پی برد. با توجه به بررسی‌های صورت گرفته در این تحقیق نشان داده شد که روش پیشنهادی ارائه شده در

این تحقیق قادر است تا با دقت بالایی نسبت به روش‌های مشهوری همچون میانگین متحرک خود همبسته [۲۷]، جستجوی نزدیک‌ترین همسایگی [۱۳]، رگرسیون خطی [۲۶] و الگوریتم پیش‌بینی سری زمانی [۲۷] عمل کند. در این روش پیشنهادی از درخت تصمیم ID3 استفاده شده است که در روند کار بهبودهای مختلفی روی این درخت اعمال شده و در نتیجه درخت تصمیم پیشنهادی ایجاد شده، دارای کم‌ترین ارتفاع ممکن و بالطبع دارای سربار محاسباتی و حافظه‌ای بسیار پایینی است. در ابتدای کار برای کاهش بیشتر سربار پردازشی از الگوریتم انتخاب با استفاده از روش ترکیبی ژنتیک و شبکه عصبی استفاده شده است. با استفاده از الگوریتم انتخاب ویژگی می‌توان از میان ویژگی‌های موجود در دیتاست، تنها ویژگی‌های با تاثیرگذاری بالا را شناسایی نمود و در نتیجه نویز را در داده‌ها از بین برد و سربار محاسباتی برای ویژگی‌های بی‌تاثیر در خروجی الگوریتم را حذف نمود. با این کار علاوه بر کاهش سربار محاسباتی و حافظه‌ای، دقت نیز به نوبه خود افزایش می‌یابد.

ارزیابی عملکرد پیش‌بینی قیمت سهام و انتخاب سبد بهینه با توجه به تغییرات قیمت آینده سهام با استفاده از روش ارایه شده در این پژوهش نشان می‌دهد که این روش قادر به پیش‌بینی نوسانات شدید با الگوهای غیرخطی بوده و پیش‌بینی‌ها رضایت‌بخش است. همچنین، بررسی عملکرد پیش‌بینی قیمت سهام بدون استفاده از روش پیشنهادی در این پژوهش نشان می‌دهد که این روش قادر است که نوسانات رخ داده در قیمت سهام را به خوبی پیش‌بینی نماید. روش حاضر دارای دقت بالایی می‌باشد و از معایب روش پیشنهادی می‌توان به سربار زمانی اشاره کرد که کمی فرایند تحلیل را زمان‌بر می‌کند، البته اگر در این حالت از ماشین‌های قوی استفاده شود، روند می‌تواند سریع‌تر نیز صورت گیرد.

ارزیابی کارایی رویکرد پیشنهادی برای پیش‌بینی روند تغییرات قیمت سهام از طریق محاسبه شاخص RMSE نشان می‌دهد که بهترین نتایج مربوط به روش پیشنهادی می‌باشد. با استفاده از رویکرد ارایه شده در این تحقیق می‌توان بهترین سبد سهام را انتخاب نمود، زیرا سهامی در سبد بهینه سهام قرار می‌گیرد که دارای بیش‌ترین تغییرات قیمت در جهت افزایش برای آینده باشد و در این حالت می‌توان به نتایج بسیار قابل قبول دست یافت.

البته از نتیجه روش بیان‌شده در این تحقیق می‌توان علاوه بر انتخاب سبد بهینه سهام، برای فروش سهامی که احتمال سقوط آنها در آینده پیش‌بینی می‌شود، نیز استفاده نمود. بنابراین می‌توان از نتیجه این تحقیق نه تنها برای انتخاب سبد بهینه سهام برای خرید، بلکه برای فروش سهام کم‌ارزش و محتمل بر ضرر نیز استفاده نمود. در حوزه مورد مطالعه در کارهای آتی، پیشنهادهای زیر مطرح می‌گردد.

- استفاده از روش مبتنی بر شبکه‌های فازی-عصبی به منظور استخراج قوانین معاملاتی سهام و مقایسه نتایج حاصل با نتایج ارایه شده در این مقاله.
- استفاده از روش‌های مدل‌سازی احتمالاتی نظیر روش مونت کارلو برای در نظر گرفتن عدم قطعیت‌هایی که ممکن است رخ دهد و روند قیمت سهام را با تغییرات شدید مواجه کند که این باعث دقت بالاتر می‌شود و می‌تواند سبد بهینه‌تری را برای سهام ایجاد نماید.

منابع

- [1] Tehrani, R., Hendijanizadeh, M., Noruzian Lakvan, E. (2015). Novel approach to fulfill active portfolio management and automatic stock trading based on feature selection algorithm. *Journal of Investment Knowledge*, 14(13), 107-126.
- [2] Raei, R., Pouyanfar, A. (2004). *Advanced investment management*. Tehran: Samt Publication.
- [3] Glasserman, P. (2013). *Monte Carlo methods in Financial Engineering*. Springer-Verlag, New York.
- [4] Sharpe, F. W. (2011). *Investment Management*. Translator (Seyed Majid Shariat Panahi). Tehran.
- [5] Farid, S. (2014). *Risk management: grouping stocks using self-organizing neural networks*. Master's thesis, Faculty of Management and Economics, Sharif University of Technology.
- [6] Alam Tabriz, A., Ali Afshari, M., Maleki, M.H., Mohammadi, J. (2007). Optimal selection of stock portfolio using neural-artificial network model in Tehran Stock Exchange. *First International Conference on Management and Innovation*, Shiraz, Iran.
- [7] Homayounfar, M., Daneshvar, A., Rahmani, J. (2015). Developing Meta-heuristic AntLion-Genetic and PBILDE Algorithms to Portfolio Optimization in Tehran Stock Exchange. *Financial engineering and portfolio management*, 9(34), 381-404.
- [8] Bertsimas, D., Shioda, R. (2009). Algorithm for cardinality constrained quadratic optimization. *Computational Optimization and Applications*, 43, 1-22.
- [9] Shafiei, H. (2015). *A Fuzzy Neural Network Model for Stock Price Prediction*, M.Sc. Thesis, Faculty of Industry, Sharif University of Technology, 2015.
- [10] Talbi, El-Ghazali. *Metaheuristics: From Design to Impelementation*, John Wiley and Sons 2009.
- [11] Xu, Y., Yang, G., Pan, C. (2013). A Heuristic Based on PSO for Irregular Cutting Stock Problem[J]. *IFAC Proceedings Volumes*, 46(13), 473-477.
- [12] Yaghini, M., Akhavan, R. (2010). DIMMA: A Design and Implementation Methodology for Metaheuristic Algorithms - A Perspective from Software Development. *International Journal of Applied Metaheuristic Computing*, 1(4), 57-74.
- [13] Angelova, M., Olympia, R., Tania, P. (2015). InterCriteria analysis of crossover and mutation rates relations in simple genetic algorithm." *Computer Science and Information Systems (FedCSIS)*, 2015 Federated Conference on. IEEE, 2015.
- [14] Adebisi, A. A., Adewumi, A. O., Ayo, C. K. (2014). Comparison of arima and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics*, 2014.
- [15] Billah, M., Waheed, S., Hanifa, A. (2015). Predicting Closing Stock Price using Artificial Neural Network and Adaptive Neuro Fuzzy Inference System ANFIS: The Case of the Dhaka Stock Exchange. *International Journal of Computer Applications*, 129(11), 1-5.
- [16] Brealey, R., Stewart, A., Myers, C., Marcus, A.J. (2016). *Fundamentals of Corporate Finance*. McGraw-Hill, New York, 2016.
- [17] Miralavi, S.H., Pourzamani, Z. (2019). Providing a model for predicting stock prices using ultra-innovative neural networks. *Financial engineering and portfolio management*, 10(40), 57-83.
- [18] Sadeghi Sharif, S.L., Farazmand, S. (2017). Assessment of Stock Price Predictions Using Artificial Neural Network (ANN) . *Quarterly Journal of Fiscal and Economic Policies*, 5(17), 97-115.
- [19] Khoshkhandeh, M., Haghghi Niat, R. (2016). Considering the effective factors in predicting the Tehran Stock Exchange price index by improving the grasshopper optimization algorithm in selecting the best examples in the neural network multiple training model. *The Third International Conference on Management, Economics and Accounting*. East Azerbaijan, Iran.
- [20] Afesharirad, M., Nazari, R. (2018). Portfolio Optimization in Selected Tehran Stock Exchange. *Regional Science Inquiry*, 4(11), 59-76.
- [21] Rostami, S. (2016). *A Study of the Efficiency of Machine Learning Algorithm in Predicting the Tehran Stock Exchange*. The Second National Conference on Modern Management Sciences and Socio-Cultural Planning in Iran. Qom, Iran.
- [22] Yazdi, H., Tavakoli, Moghadam, R., Golriz, B. (2015). Comparison of the performance of four new meta-innovative algorithms for solving mathematical optimization problems. *The 8th International Conference of the Iranian Association for Operations Research*, Ferdowsi University of Mashhad. Mashhad, Iran.
- [23] Chen Y., Mabu, S., Hirasawa, K. (2011). Genetic relation algorithm whit guided mution for the large -scale portfolio optimization" *Expert System Whit Applications*. *Expert Systems with Applications*, 4(38), 3353-3363.

- [24] Easton, P.D. (2014). PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The accounting review*, 1(79), 73-95.
- [25] Quandl Package. (2020). Free datasets. Retrived from: <https://www.Quandl.com>
- [26] Miranda, B., Vinicius, H., Sobreiro, A., Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, 4(3), 183-201.
- [27] Afeef, M., Ihsan, A., Zada, H. (2018). Forecasting Stock Prices through Univariate ARIMA Modeling. *NUML International Journal of Business & Management*, 13(2), 130-142.