مجله مهندسی مالی و مدیریت اوراق بهادار
شماره سی‌وپنجم / تابستان ۱۳۹۷

پیش بینی روند قیمت در بازار سهام با استفاده از الگوریتم جنگل تصادفی

الهام غلامیان[۱]
سیدمحمدرضا داودی[۲]

تاریخ دریافت: ۹۶/۰۸/۰۲ تاریخ پذیرش: ۹۶/۰۹/۲۰

چکیده

فعالان بورس درصدد دستیابی و به کارگیری روش‌هایی هستند تا بتوانند با پیش‌بینی آتی قیمت سهام، سود سرمایه خود را افزایش دهند. بنابراین، ضروری به نظر می‌رسد که روش‌های مناسب، صحیح و متکی به اصول علمی در تعیین قیمت آینده سهام فرآروی افراد سرمایه‌گذار قرار گیرد. تاکنون روش‌های مختلفی جهت نیل به این هدف معرفی شده‌اند که اغلب روش‌های آماری و هوش مصنوعی هستند. در پژوهش حاضر با استفاده از رویکرد جنگل تصادفی که در زمره روش‌های طبقه‌بندی هوش مصنوعی می‌باشد، به همراه شاخص‌های فنی: شاخص قدرت نسبی قیمت، استوکاستیک، حجم تعادل موازنه شده، ویلیامز %R، بازده‌ی روزانه و شاخص سری مک‌دی به دنبال پیش‌بینی روند قیمت در بازار سهام و مقایسه آن با روش‌های موجود است. نتیجه‌ی پژوهش بر روی داده‌های روزانه شاخص بورس اوراق بهادار تهران در سالهای ۱۳۹۳ تا ۱۳۹۵ نشان می‌دهد که دقت روش پیشنهادی در برآورد روند بازار ۶۴ درصد می‌باشد و نسبت به دو روش مقایسه شده رگرسیون لجستیک و روش کاملا تصادفی از دقت بالاتری برخوردار است.

واژه‌های کلیدی: الگوریتم جنگل تصادفی، پیش‌بینی قیمت سهام، آنتروپی، شاخص‌های تکنیکی، رگرسیون لجستیک.

۱- کارشناسی ارشد مهندسی صنایع ،گرایش سیستم‌های اقتصادی-اجتماعی، واحد دهاقان، دانشگاه آزاد اسلامی، دهاقان، ایران
elhamgh35@yahoo.com
۲- استادیار و عضو هیات علمی

۱- مقدمه

هر تصمیم‌گیری آگاهانه نیاز به پیش‌بینی دارد. در یک تعریف کلی گمانه‌زنی در مورد پیش‌بینی عبارت است از: برآورد احتمالی وقایع آینده بر اساس اطلاعات حال و گذشته.(زراء نژا، ۱۳۸۸). روند پیش‌بینی قیمت‌ها در بازار سهام به دلیل ماهیت پیچیده و پویای آن، سال‌ها مورد توجه محققان بوده است. همچنین به دلیل ابهامات و متغیرهایی که در تغییر شاخص بازار در یک روز خاص موثر است؛ یک کار بسیار بحث برانگیز است. از جمله‌ی متغیرهای تاثیرگذار بر قیمت سهام می‌توان، شرایط اقتصادی، احساسات سرمایه‌گذاران نسبت به یک شرکت خاص، رویدادهای سیاسی و غیره را نام برد؛ بنابراین بازارهای سهام به تغییرات سریع، حساس بوده و درنتیجه باعث نوسانات تصادفی در قیمت آنها می‌شود. با توجه به ماهیت بی‌نظمی و بی‌ثباتی زیاد رفتار سهام، سرمایه-گذاری در این بازار با ریسک بالایی همراه است که به منظور حداقل‌سازی این مخاطرات، دانشی که حرکت قیمت سهام در آینده را با دقت بالایی پیش‌بینی کند، نیاز است. (خایدم[1]و همکاران، ۲۰۱۶).

فعالان بوس درصدد دستیابی و به کارگیری روش‌هایی هستند تا بتوانند با پیش‌بینی آتی قیمت سهام، سود سرمایه خود را افزایش دهند. بنابراین، ضروری به نظر می‌رسد که روش‌های مناسب، صحیح و متکی به اصول علمی در تعیین قیمت آینده سهام فرآروی افراد سرمایه‌گذار قرار گیرد. اقتصاددانان برای پیش‌بینی در بیشتر موارد از روش‌های اقتصاد سنجی استفاده می‌نمایند. در این بین، از پرکاربردترین روش، آریما[2](۱) و آریما(۲) فرآیندهای خطی رگرسیونی در پیش‌بینی محسوب می‌شوند. در سال‌های اخیر به موازات پیشرفت‌های قابل توجه در پردازش سریع اطلاعات به وسیله ماشین‌های الکترونیکی، به کارگیری مدل‌های غیرخطی در میان اقتصاددانان به طور چشم‌گیری افزایش یافته است.(بدیعی و همکاران، ۱۳۹۶). از مهمترین مدل‌های غیرخطی که در سال‌های اخیر در بازارهای مالی بسیار مورد استفاده قرار گرفته و به نتایج مطلوبی نیز دست یافته است، مدل شبکه‌های عصبی مصنوعی، الگوریتم ژنتیک و الگوریتم جنگل‌تصادفی که تعمیم یافته الگوریتم درخت تصمیم می‌باشد، می‌توان نام برد. هدف این پژوهش بکارگیری الگوریتم جنگل تصادفی در پیش‌بینی قیمت سهام، و بررسی دقت آن است.

۲- مبانی نظری و مروری بر پیشینه پژوهش

در سال‌های اخیر در منابع مختلف بیان شده است که بازده سهام رابطه‌ای غیرخطی با متغیرهای مالی و اقتصادی دارد و تکنیک‌های هوش‌مصنوعی ابزار بسیار مناسبی برای حل مسائل غیرخطی می‌باشند. در برخی از منابع، شاخص بازار را برای انجام پیش‌بینی مورد استفاده قرار

داده‌اند. یکی از دلایل این کار این نکته است که شاخص بازار نماد کاملی از تمام شرکت‌های موجود در بازار است و کارایی بازار بر اساس آن تعیین می‌شود و دید کلی‌تری را به سرمایه‌گذارن می‌دهد و نمایانگر روند کلی بازار می‌باشد. (حقیقت منفرد و همکاران، ۱۳۹۱).

انتخاب ورودی‌های درست یکی از عوامل اصلی تاثیرگذار در کارآیی مدل پیش‌بینی می‌باشد. الگوریتم جنگل‌تصادفی در صورتی قادر به کشف درست اطلاعات پنهان است، که ورودی‌های مناسبی انتخاب گردد. ورودی‌های مورد مطالعه در این الگوریتم شاخص‌های تکنیکی می‌باشند. یک شاخص تکنیکی تابعی از زمان بازار است که برای سری داده‌های قیمت و یا حجم یک سهم یا یک بازار به کار گرفته شود. این روش بیانگر روش تحلیل کمی قیمت یا حجم در طی سهام زمان است که معمولا به شکل یک معادله‌ی ساده‌ی ریاضی برای نشان دادن ویژگی‌های خاص می‌باشد و علائمی را برای کمک به پیش‌بینی حرکت بازار ارائه می‌کند. هر شاخص تکنیکی نمایانگر یک دیدگاه خاص است که به وسیله‌ی آن می‌توان عملکرد بازار را تحلیل کرد و تغییرات قیمت را پیش‌بینی کرد. هیچ شاخصی همیشه درست نیست و شاخص‌های تکنیکی مختلف ممکن است بیشتر مواقع با هم در تضاد باشند. استفاده از شاخص‌های تکنیکی بسیار متفاوت است این شاخص-ها پر شمار بوده و پویایی قابل ملاحظه‌ای دارند. (محمدی،۱۳۸۳).

تجزیه و تحلیل تکنیکال، شامل تکنیک‌های پیش‌بینی است که با اندازه‌گیری الگوهای تاریخی رفتار قیمت سهام و ویژگی تاریخی سایر اطلاعات مالی به دست آمده است. تحلیلگر پس از بررسی عملکرد رفتار گذشته اطلاعات جاری مربوط به قیمت سهام را مورد بررسی قرار می‌دهد تا دریابد که آیا هیچ گونه الگوی برقرار شده‌ای قابل اعمال است یا خیر؟ اگر چنین باشد، پیش‌بینی‌هایی می‌توان انجام داد. ایده اصلی تجزیه و تحلیل تکنیکی این است که روند تغییرات سهام به وسیله تغییرات نگرش سرمایه‌گذاران که خود متاثر از عوامل متعددی است شکل می‌گیرد. با استفاده از قیمت و حجم معاملات، تحلیلگران تکنیکی به پیش‌بینی تغییرات آینده قیمت‌ها اقدام می‌کنند. تحلیلگران تکنیکی بر این عقیده‌اند که تاریخ خود را تکرار می‌کند و تغییرات آتی قیمت سهام می-تواند با توجه به قیمت‌های پیشین سهام تعیین شوند. تکنیکال‌ها فرض را بر آن می‌گذارند که عوامل بنیادی تاثیر خود را بر روی قیمت‌ها گذاشته‌اند و لذا تحلیل نمودار قیمت می‌تواند راهگشای پیش‌بینی آینده آن باشد. (کامروافر، هاشمی،۱۳۹۶).

درخت خوشه‌بندی و رگرسیون توسط بریمن[۳]، و همکاران (۱۹۸۴) آغاز شد و یک الگوریتم درخت تصمیم دودویی مستعد فرآیند متوالی یا پیش‌بینی کننده مطلق با متغیرهای هدف است و به طور بازگشتی کار می‌کند. و الگوریتم جنگل تصادفی در سال ۲۰۰۱ توسط لیو بریمان در دانشگاه اکسفورد ساخته شده است.

درخت تصمیم‌گیری یکی از الگوریتم‌های یادگیری نظارت شده غیر پارامتریک می‌باشد که به صورت عمده در مسایل طبقه‌بندی کاربرد دارد. در این روش کلاس یا طبقه یک نمونه براساس فرآیند شاخه‌بندی مشخص می‌شود. درخت‌های تصمیم‌گیری را می‌توان به درخت‌های تصمیم-گیری با خروجی کلاس و درخت‌های تصمیم پیوسته طبقه‌بندی کرد. درخت رگرسیونی در مورد متغیرهای پیوسته به کار می‌روند، در حالی که درخت‌های طبقه‌بندی در مورد داده‌های دارای کلاس یا رتبه به کار می‌روند. در مورد درخت تصمیم‌گیری مفهومی به نام گره ریشه وجود دارد که همان جمعیت اولیه است که باید به صورت هم‌گون به دو گروه تقسیم شود. گره‌های پایانی را برگ یا ترمینال می‌نامند.

یکی از مفاهیمی که در مورد درخت تصمیم‌گیری باید مورد توجه قرار گیرد مساله اورفیتینگ[۴] می‌باشد. این مساله به این مطلب اشاره دارد که در صورتی که تقسیم بندی شاخه‌های درخت زیاد باشد، درخت تصمیم‌گیری با موفقیت ۱۰۰٪ جداسازی کلاس‌ها را بدون یادگیری انجام می‌دهد. درواقع درخت تصمیم‌گیری چیزی یاد نمی‌گیرد بلکه الگوها را حفظ می‌کند. برای جلوگیری از این رویداد می‌توان محدودیت‌هایی مانند عمق درخت، حداکثر تعداد گره‌های ترمینال و حداقل نمونه لازم برای جداسازی در گره‌های پایانی را مد نظر قرار داد. مانند بسیاری از مفاهیم و سیستم‌ها در هوش مصنوعی، درخت تصمیم نیز از دو عدم توازن ،در دو مفهوم بایاس[۵] و واریانس رنج می‌برد. مفهوم بایاس اشاره به این دارد که متوسط خروجی سیستم چه مقدار با مقدار مورد واقعی مورد انتظار فاصله دارد و مفهوم واریانس اشاره به آن دارد که با تعویض نمونه یا استفاده از داده‌های برون نمونه‌ای دقت سیستم به چه میزانی فرق می‌کند. اگر یک درخت با عمق کم انتخاب کنیم، از خطای بایاس رنج می‌برد و واریانس را کاهش می‌دهد و اگر از یک درخت با عمق زیاد استفاده کنیم از واریانس بالاتری برخوردار می‌شود، در حالی که بایاس را کاهش می‌دهد. بنابراین همواره باید درخت تصمیم به گونه‌ای انتخاب شود که یک توازن بین این دو برقرار شود.

روش بگینگ[۶] برای کاهش واریانس استفاده می‌شود و نتایج حاصل از پیش‌بینی چند سیستم پیش‌بینی کننده که هر کدام با استفاده از زیر نمونه‌هایی (با امکان تکرار در نمونه‌گیری) آموزش دیده‌اند را بر اساس معیارهایی مانند میانگین،میانه و... ترکیب می‌کند.

از روش‌های پیاده‌سازی بگینگ در مورد درخت تصمیم، رویکرد جنگل تصادفی می‌باشد. روش جنگل تصادفی جزئی از روش‌های طبقه‌بندی تحت نظارت گروهی در حوزه داده‌کاوی است. جنگل تصادفی یا به اختصار RF ،پتانسیل بسیار زیادی برای تبدیل به یک روش محبوب برای طبقه‌بندی-های آینده دارد زیرا کارایی آن قابل مقایسه با دیگر روش‌های گروهی می‌باشد. به عنوان یک الگوریتم گروهی(رأی گیری[۷])، RF چندین درخت تصمیم مختلف به عنوان طبقه‌بندی‌های پایه

تولید می‌کند و رأی اکثریت را برای ترکیب با نتایج درختان اصلی اعمال می‌کند. (سعادت،فرساد،۱۳۹۵).جنگل تصادفی هم توانایی رگرسیون، و هم توانایی کلاس‌بندی را با هم دارد. روش کار در یک جنگل تصادفی چنین است که، داده‌های نمونه‌ای با عمل جایگذاری به تعداد n مجموعه نمونه تقسیم می‌شود و از هر دسته، نمونه‌ای برای آموزش یک درخت استفاده می شود. تمام درخت‌ها عمق مشخصی دارند و در هر گره تقسیم یک ویژگی یا فیچر به تصادف از بین مجموعه ویژگی‌ها انتخاب می‌شود و تقسیم یا شاخه‌بندی بر اساس آن صورت می‌گیرد. به علت استفاده از چندین دسته نمونه این روش مشکل داده‌های پرت و داده‌های گم شده را ندارد.

در سال‌های اخیر تحقیقات زیادی در زمینه‌های پیش‌بینی سهام با استفاده از تکنیک‌های هوش مصنوعی و داده‌کاوی صورت پذیرفته است که در ادامه به اختصار توضیح داده می‌شود.

بدیعی و همکاران،(۱۳۹۶) در تحقیقی تحت عنوان " پیش‌بینی رفتار قیمتی سهام در بورس اوراق بهادار با استفاده از شبکه‌ی عصبی مصنوعی (مطالعه موردی شرکت پالایش نفت اصفهان)"به پیش‌بینی قیمت در بازار سهام با استفاده از تکنیک‌های هوش‌مصنوعی پرداخته است که نتایج بدست آمده نشان می‌دهد، مدل شبکه عصبی دارای خطای پایین و قدرت توضیح دهندگی بالا و در نتیجه از قدرت پیش‌بینی خوبی برخوردار می‌باشد.

تقی‌پوریان و همکاران (۱۳۹۴) در پژوهشی تحت عنوان "ارزیابی مدل‌های درخت تصادفی در پیش‌بینی عملکرد مالی شرکت‌های پذیرفته‌شده در بورس اوراق‌بهادار تهران" که در آن نمونه‌ای متشکل از ۲۱ نسبت در ۵۳۴ شرکت پذیرفته‌شده در بورس اوراق‌بهادار تهران در فاصله بین سال-های ۱۳۹۰ تا ۱۳۹۳ به عنوان متغیرهای مستقل و دو نسبت بازده دارایی‌ها وبازده حقوق صاحبان سهام به عنوان متغیرهای وابسته انتخاب شده است. نتایج تحقیق حاکی از آن است که بین دو متغیر بازده دارایی‌ها و بازده حقوق صاحبان سهام، بازده حقوق صاحبان سهام از لحاظ ارزیابی‌های به دست آمده از صحت بالاتری برخوردار است و در بین چهار درخت تصمیم سی فایو از بهترین شاخصه‌های ارزیابی برخوردار بود.

شهریاری و همکاران (۱۳۹۳) در مقاله‌ای تحت عنوان " یادگیری درخت تصمیم‌گیری " که نتایجی به این شرح دارد با توجه به حجم بالای اطلاعات در بانک‌های داده و یافتن اطلاعات مفید و مناسب در بانک‌های اطلاعاتی ضرورت وجود ابزار و روشی برای پردازش و طبقه‌بندی داده‌های ذخیره شده و نمایش اطلاعات مفید وجود دارد. یادگیری درخت تصمیم یکی از پرکاربردترین وکارآمدترین متدهای یادگیری استقرایی برای پردازش داده‌ها می‌باشد. این متد یادگیری برای توابع گسسته مقادیر باداده‌های خطادار به کار می‌رود. درخت تصمیم‌گیری برخلاف روش‌هایی مانند شبکه‌های عصبی که چگونگی عملکرد خود را ارائه نمی‌کنند، پیش‌بینی خود را در قالب

قوانینی که از نظر پارامترهای آماری برازش مناسبی ارایه می‌کند. درختان تصمیم قادر به تولید توصیفات قابل درک برای انسان، از روابط موجود در یک مجموعه داده‌ای هستند و می‌توانند برای وظایف دسته‌بندی وپیش‌بینی بکار روند. هدف از این مطالعه نشان دادن روش یادگیری درخت تصمیم‌گیری به عنوان روشی مناسب و قوی برای داده‌هایی با حجم بالا و یا ناقص بوده است. روشی که به موفقیت قابل توجه‌ایی در سیستم‌های تجاری دست یافته است.

مهرآرا و همکاران (۱۳۹۱) در تحقیقی تحت عنوان "تکنیک‌های داده‌کاوی در پیش‌بینی قیمت سهام بورس اوراق‌بهادار"به تحقیق درباره‌ی تکنیک‌های داده‌کاوی و پیش‌بینی قیمت سهام پرداخته. که از جمله‌ی تکنیک‌های داده‌کاوی می‌توان به درخت تصمیم‌گیری اشاره کرد.

البرزی و همکاران (۱۳۹۱) در تحقیقی تحت عنوان " بکارگیری تکنیک‌های خوشه‌بندی و الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک‌ها" در این پژوهش پنج درخت تصمیم‌گیری مبتنی برالگوریتم C4.5 در هر خوشه با مجموعه ویژگی‌های منتخب ساخته می‌شود. بهترین درختان تصمیم‌گیری در هر خوشه مبتنی بر معیارهای بهینگی مورد نظر در این مقاله انتخاب شده و با هم ترکیب می شوند تا درخت تصمیم‌گیری نهایی برای رسیدن به نتایج بکار GATree برای اعتبارسنجی مشتریان بانک ایجاد شود. ابزار یادگیری ماشین وکا و نرم‌افزار گرفته شده است. نتایج پژوهش نشان می‌دهد که استفاده از مدل ترکیبی پیشنهادی در ساخت درخت تصمیم‌گیری منجر به افزایش دقت طبقه‌بندی نسبت به بسیاری از الگوریتم‌های مقایسه شده می‌شود؛ ولی پیچیدگی الگوریتم مدل ترکیبی پیشنهادی از برخی الگوریتم‌های طبقه-بندی مقایسه شده بیشتر است.

فروغی و منادجمی،(۱۳۹۰) در تحقیقی تحت عنوان "اجرای فن درخت تصمیم برای پیش‌بینی ورشکستگی " از درخت تصمیم C&R برای پیش‌بینی ورشکستگی مالی شرکت‌های پذیرفته شده در بورس اوراق‌بهادار تهران بین سال‌های ۱۳۸۷تا ۱۳۸۸ استفاده کرده است. متغیرهای پیش‌بینی کننده شامل۲۵ نسبت مالی به دست آمده از صورت‌های مالی است که در۴ گروه نسبت‌های نقدینگی، اهرمی، کارایی و سوددهی بخش‌بندی شده است. دقت آموزش مدل ۹۴/۵ درصد و دقت تست آن ۹۹ درصد گزارش شد.

آهنگری (۱۳۹۰) در مقاله‌ای تحت عنوان "بکارگیری الگوریتم درخت تصمیم جهت پیش‌بینی شرکت‌های ورشکسته و غیرورشکسته پذیرفته شده در بورس اوراق‌بهادار" در این تحقیق با استفاده از داده‌های نسبت‌های مالی ۱۴۴ شرکت پدیرفته شده در بورس تهران در بازه زمانی انتخابی نمونه شامل سال‌های ۱۳۸۴ تا ۱۳۸۸ فرآیند داده‌کاوی صورت پذیرفت. الگوریتم‌های مختلف درخت تصمیم QUEST.CHAID.C5. CART بر روی داده‌های نهایی اجرا و نتایج پیش‌بینی این

الگوریتم‌های طبقه‌بندی مورد مقایسه قرار گرفتند. نتایج نشان‌دهنده‌ی برتری الگوریتم CART با صحت متوسط ۹۴,۴۳ درصد در هریک از سال‌هایT ، T-1 ، T-2 است.

ژینجی[۸] و همکاران. (۲۰۱۴) در مقاله‌ای تحت عنوان " پیش‌بینی روند سهام با شاخص‌های فنی با استفاده از SVM" با بررسی شاخص‌های فنی مختلف از جمله RSI، در تعادل دوره، ویلیامز٪ Rو غیره .. به عنوان ویژگی‌های انتخاب شده همچنین از سهام سه شرکت (AAPL، مایکروسافت، AMZN) در دوره زمانی (۲۰۱۰/۰۱/۰۴) تا ۱(۲۰۱۴/۱۲/۱۰) استفاده شده و از بین ۸۴ مورد ویژگی‌ها یک درخت بسیار تصادفی اجرا شد و سپس با استفاده از SVM دادهای آموزشی تغذیه شدند.

دای و ژانگ[۹] (۲۰۱۳) در تحقیقی تحت عنوان " یادگیری ماشین در روند پیش‌بینی قیمت سهام" نتایجی به شرح زیر همراه داشته. داده‌های آموزشی مورد استفاده در تحقیقات خود، اطلاعات سهام 3M بود که داده‌ها شامل اطلاعات روزانه ۱۴۷۱ سهام از تاریخ (۲۰۰۸/۰۱/۰۹) تا (۲۰۱۳/۱۱/۰۸)، الگوریتم‌های مختلف برای آموزش داده‌ها در سیستم پیش‌بینی انتخاب شدند. این الگوریتم‌ها، رگرسیون لجستیک، آنالیز تفکیک درجه دوم، وSVM بوده که برای پیش‌بینی نتیجه قیمت سهام و دریک روز آینده و در بلندمدت برای N روز آینده استفاده شد. مدل پیش‌بینی روز بعد نتایج دقت اعم از۴۴/۵۲ درصد به۵۸/۲ درصد را به همراه داشت.

ایودل ادبی[۱۰]ا (۲۰۱۱) یک روش هیبریداسیون که ترکیبی از متغیرهای از تجزیه و تحلیل فنی و بنیادی شاخص های بازار سهام را برای پیش بینی قیمت سهام در آینده به منظور بهبود در روش‌های موجود استفاده کرد. تمرکز این مقاله به منظور بهبود دقت پیش بینی قیمت سهام استفاده از روش ترکیبی که ترکیبی از متغیرها از تجزیه و تحلیل فنی و بنیادی برای ایجاد مدل شبکه عصبی برای پیش بینی قیمت سهام بود. متغیرهای تجزیه و تحلیل فنی در بررسی او، هسته اصلی شاخص های بازار سهام (قیمت سهام در حال حاضر، قیمت باز شدن، بسته شدن قیمت، حجم، بالاترین قیمت و پایین ترین قیمت و غیره) در حالی که متغیرهای اساسی تجزیه و تحلیل شاخص های عملکرد شرکت (قیمت در سود سالانه، شایعه / اخبار، ارزش دفتری و وضعیت مالی وغیره)

یو و ونجوان[۱۱] (۲۰۱۰) برای بررسی اینکه کدامیک از نسبت های مالی قوی‌ترین تاثیر را بر رشد سود شرکت های تدارکاتی دارند، از روش C5.0 که یکی از روش‌های درخت تصمیم است استفاده کردند و به بررسی تحلیل‌های مالی از جمله تجزیه وتحلیل سوددهی پرداختند و نتایج نشان داد روش C5.0 دارای دقت۹۵ درصد است.

ونگ[۱۲] و همکاران (۲۰۰۹) با استفاده از درخت تصمیم و ۵ نسبت مالی به پیش‌بینی سود سال آتی پرداختند. آنها نشان دادند بین مدل درخت تصمیم استاندارد و درخت تصمیم افزایشی، درخت تصمیم کیسه‌ای بهترین عملکرد را برای پیش‌بینی بازده سهام دارد.

فینگ، (۱۹۹۴) به بررسی سود در پیش‌بینی سودها و جریان‌های نقدی آتی پرداخت. در این تحقیق که دوره زمانی سالهای (۱۹۳۵ – ۱۹۸۷)را پوشش می‌داد و نمونه‌ای متشکل از ۵۰ شرکت پذیرفته شده در بورس نیویورک را در بر می‌گرفت، این نتیجه حاصل شد که برای ۸۸ درصد شرکت‌های نمونه، سودهای گذشته را می‌توان به عنوان یک پیش‌بینی کننده خوب برای سودهای آتی دانست.

۳- روش شناسی پژوهش

۳-۱- نوع پژوهش، نمونه و جامعه آماری

تحقیق حاضر از بعد هدف کاربردی و از نظر نوع روش در زمره تحقیقات پیمایشی می‌باشد. و از این جهت که گردآوری دادها از طریق نمونه‌گیری از جامعه، تحقیق میدانی و تجزیه و تحلیل پاسخ های جامعه به سولات پرسشنامه صورت می‌گیرد، از نوع پیمایشی یا زمینه‌یابی است. در این تحقیق با استفاده از روش جنگل تصادفی به آنالیز داده‌ها خواهیم پرداخت و خروجی مربوط به این روش را در برنامه متلب[۱۳] استخراج می‌نماییم. نمونه و جامعه‌ی آماری پژوهش شامل ۷۲۳ داده از شاخص کل بورس اوراق بهادار تهران طی یک دوره‌ی سه ساله (۱۳۹۳-۱۳۹۵) می‌باشد. داده‌های شاخص به صورت روزانه بوده و از نرم‌افزار ره‌آوردنوین استخراج شده است.

۳-۲- سوال های پژوهش

۱) چگونه با بهره‌گیری از الگوریتم درخت تصادفی می‌توان قیمت سهام را پیش‌بینی کرد؟

۲) توانایی الگوریتم جنگل تصادفی در مقایسه با روش رگرسیون لجستیک و روش کاملا تصادفی (پرتاب تاس) چگونه است؟

۳-۳- متغیرهای پژوهش

فرض کنیم داده‌های روزانه قیمت شاخص به صورت سری زمانی $\{x_i\}_{i=1}^{N}$ باشند. در ابتدا به بررسی نحوه محاسبه شش اندیکاتور برای شاخص، به‌عنوان ویژگی یا فیچر درخت تصمیم‌گیری پرداخته می‌شود. از آنجا که درخت تصمیم در فریم روزانه مورد استفاده قرار خواهد گرفت و هدف

آن پیش‌بینی جهت بازار در روز آتی می‌باشد، تمام اندیکاتورها نیز در فریم روزانه به صورت زیر محاسبه می‌گردند:

- **اندیکاتور مک دی[۱۴]**

برای محاسبه اندیکاتور مک‌دی ابتدا میانگین متحرک نمایی برای دوره‌های زمانی ۱۲و ۲۶ روزه محاسبه می‌شود و سپس از تفاضل آنها سری مک‌دی محاسبه می‌شود. روابط(۱)،(۲)،(۳).

رابطه ۱)
$$macd_i = ema^{12}_i - macd^{26}_i \qquad i \geq 26$$

رابطه ۲)
$$ema^n_n = \frac{x(1)+x(2)+\ldots+x(n)}{n}$$

رابطه ۳)
$$ema^n_i = \frac{2}{1+n}\, x(i) + (1-\frac{2}{1+n}) ema^n_{i-1} \qquad i > n$$

- **اندیکاتور RSI[۱۵] یا شاخص قدرت نسبی**

این اندیکاتور همواره بین دو سطح صفر و ۱۰۰ در حال نوسان می‌باشد. در تنظیمات این اندیکاتور عموما از تنظیم ۱۴ روزه استفاده می‌شود و فرمول محاسبه این اندیکاتور به صورت روابط (۴) و(۵) محاسبه می‌شود.

رابطه ۴)
$$RSI = 100 - \frac{100}{1+RS}$$

رابطه ۵)
$$\mathrm{RS} = \frac{\text{Average Gain Over past 14 days}}{\text{Average Loss Over past 14 days}}$$

در فرمول RS صورت کسر برابر مجموع تمام افزایش قیمت‌ها در ۱۴ روز گذشته و مخرج برابر مجموع تمام کاهش قیمت ها در ۱۴ روز گذشته می‌باشد که البته یک عدد مثبت است.

- **اندیکاتور so[۱۶]**

در محاسبه این نوسانگر دو مقدار K% و D% به صورت روابط (۶) و(۷) محاسبه می‌شوند.

رابطه ۶)
$$\%K_i = \frac{\left(x(i) - \min\{x(i),x(i-1),\ldots,x(i-14+1)\}\right)}{\left(\max\{x(i),x(i-1),\ldots,x(i-14+1)\} - \min\{x(i),x(i-1),\ldots,x(i-14+1)\}\right)} \times 100$$

رابطه ۷)
$$\%D_i = \frac{\%K_i + \%K_{i-1} + \%K_{i-2}}{3} \qquad i \geq 3$$

بنابراین هر چه K% بزرگتر شود، یعنی فاصله بیشتری بین ماکسیمم و مینیمم قیمت توسط قیمت‌های جدید پو شش داده می‌شود. و نزدیکی این عدد به ۱۰۰ نشان از پتانسیل بیشتر سهم برای ریزش قیمت دارد. از طرفی هر چه K%کمتر شود یعنی فاصله کمتری بین ماکسیمم و مینیمم قیمت توسط قیمت‌های جدید پوشش داده می‌شود و نزدیکی این عدد به صفر نشان از پتانسیل بیشتر سهم برای ریزش قیمت دارد.

- **شاخص درصد آر ویلیامز[۱۷]**

این شاخص به مانند اندیکاتور SO می‌باشد با این تفاوت که معیار فاصله را نسبت به بیشترین قیمت در یک دوره ۱۴ روزه می‌سنجد. این اندیکاتور بین صفر تا ۱۰۰- نوسان می‌کند. نحوه محاسبه این اندیکاتور به صورت رابطه‌ی (۸) می‌باشد:

رابطه ۸)

$$\%R_i = \frac{\left(\max\{x(i),x(i-1),...,x(i-14+1)\} - x(i)\right)}{\left(\max\{x(i),x(i-1),...,x(i-14+1)\} - \min\{x(i),x(i-1),...,x(i-14+1)\}\right)} \times -100$$

- **اندیکاتور OBV[۱۸] یا شاخص حجم متوازن**

این شاخص به تعداد سهام‌های معامله شده در یک روز می‌پردازد. بصورتیکه اگر تعداد سهام معامله شده در یک روز نسبت به روز قبل افزایش یابد، نشان دهنده افزایش حجم خرید می‌باشد و با رابطه‌ی (۹) محاسبه می‌شود:

$$OBV = OBV_{\text{دیروز}} + OBV_{\text{امروز}} \qquad (۹)$$

اگر تعداد سهام معامله شده یک روز نسبت به روز قبل کاهش یابد نشان دهنده افزایش حجم فروش می‌باشد و با رابطه‌ی (۱۰) محاسبه می‌گردد:

$$OBV = OBV_{\text{دیروز}} - OBV_{\text{امروز}} \qquad (۱۰)$$

و اگر تعداد سهام معامله شده یک روز نسبت به روز قبل تغییر نکند نشان دهنده ثبات حجم خرید و فروش می‌باشد که با رابطه‌ی (۱۱) بدست می‌آید:

$$OBV = OBV_{\text{دیروز}} \qquad (۱۱)$$

۳-۴- مدل پژوهش

پس از بررسی متغیرها به طراحی درخت تصمیم پرداخته می‌شود. خروجی درخت تصمیم باینری می‌باشد، به صورتی که عدد یک نشان دهنده روند صعودی شاخص برای روز آتی، و عدد دو نشان دهنده روند نزولی شاخص برای روز آتی می‌باشد. از جمله پارامترهایی که در جنگل تصادفی باید مشخص گردد، تعداد درخت‌ها و عمق آنها می‌باشد. پارامتر دیگری که در ساخت هر درخت جنگل تصادفی باید مورد انتخاب قرار گیرد، تعداد داده‌های آموزشی و مجموعه فیچرهای مورد استفاده می‌باشد. در واقع هر درخت تصمیم از تمام داده‌های آموزشی و از تمام فیچرها استفاده نمی‌کند، و به صورت تصادفی از تعدادی از آنها استفاده می‌کند. بنابراین در صورتی که $\{x_1, x_2, \ldots, x_n\}$ داده‌های آموزشی باشند (که در پژوهش حاضر داده‌های روزانه شاخص هستند)، به صورت تصادفی زیر مجموعه‌ای از آنها مانند $\{y_1, y_2, \ldots, y_m\}$ انتخاب می‌شود. که البته تعداد m به گونه‌ای انتخاب می‌شود که$m> 0/66n$. باشد. همچنین اگر تعداد فیچرهای مورد استفاده برابر k فیچر باشد،هر درخت به صورت تصادفی و با تکرار از $h \leq k$ فیچر استفاده می‌کند.

در ادامه نحوه ساخت هر درخت، با بررسی یک گره شرح داده می‌شود. فرض کنیم مجموعه داده‌های آموزشی برای یک درخت تصمیم برابر $\{y_1, y_2, \ldots, y_m\}$ و تعداد فیچرهای مورد استفاده برابر h باشد. در گره شماره یک، یکی از فیچرها به صورت تصادفی انتخاب می‌شود و فرآیند شاخه‌بندی بر اساس آن فیچر صورت می‌گیرد. جمعیت وارد شده به گره بر حسب نقطه برش به دو جمعیت تقسیم می‌شود.

نقطه برش p به صورت مستقیم محاسبه می شود. برای این منظور فاصله کمترین و بیشترین مقادیر فیچری که به صورت تصادفی برای گره در نظر گرفته شده است به ۱۰۰ قسمت مساوی تقسیم می‌شود و از بین تقسیمات صورت گرفته، مقدار نقطه برش برابر عددی می‌باشد که کمترین مقدار آنتروپی را ایجاد کند. نحوه محاسبه آنتروپی در ادامه بررسی می‌شود. فرض کنیم p یک نقطه کاندید برای برش باشد. این نقطه جمعیت وارد شده به گره را به دو جمعیت:

$$A = \{y \leq p\}$$
$$B = \{y > p\}$$

تقسیم می‌کند. در صورتی که تعداد اعضای مجموعه A برابر a_1 و تعداد اعضای A که نشان دهنده روند صعودی هستند، a_2 باشد؛ و همچنین تعداد اعضای مجموعه B برابرb_1 و تعداد اعضای B که نشان دهنده روند صعودی هستند، b_2 باشد؛ مقدار آنتروپی متناظر با نقطه برش p برابر است:

رابطه ۱۲)

$$H = \frac{a_1}{m}(-\frac{a_2}{a_1}\log(\frac{a_2}{a_1}) - \frac{(a_1-a_2)}{a_1}\log(\frac{(a_1-a_2)}{a_1})) + \frac{b_1}{m}(-\frac{b_2}{b_1}\log(\frac{b_2}{b_1}) - \frac{(b_1-b_2)}{b_1}\log(\frac{(b_1-b_2)}{b_1}))$$

خروجی گره‌های پایانی نیز برابر بیشترین رای گره انتخاب می‌شود. بنابراین در صورتی که اکثریت جمعیت وارد شده به یک گره پایانی دارای روند صعودی باشند، خروجی گره برابر یک؛ و در غیر این صورت برابر دو می‌باشد. خروجی درخت تصمیم نیز به صورت مشابه برابر بیشترین رای صادر شده از گره‌های پایانی انتخاب می‌شود و در نهایت خروجی جنگل تصادفی برابر بیشترین رای صادر شده از درخت‌های تصمیم می‌باشد. در نهایت داده‌های تست یکی یکی وارد جنگل تصادفی شده و خروجی آنها مشخص می‌گردد و با خروجی واقعی مقایسه می‌گردد تا دقت جنگل تصادفی تعیین شود.

۴- یافته‌های پژوهش

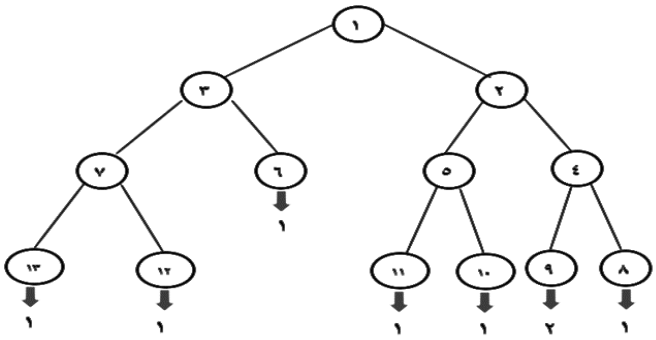
داده‌های روزانه شاخص کل در بازه ۱۳۹۵-۱۳۹۳شامل ۷۲۳ داده است. که از ۵۰۰ داده اول به عنوان داده‌های آموزشی و از ۲۰۰ داده آخر به عنوان داده‌های تست در درخت تصمیم استفاده خواهد شد.

لیست فیچر یا ویژگی‌های مورد استفاده در جنگل تصادفی همراه آماره آنها در جدول ۱ ارایه شده است:

جدول ۱- لیست فیچرهای و آماره ی آنها

شماره	فیچر
۱	PROC::بازده روزانه
۲	MACD:اندیکاتور مک دی
۳	SO:اندیکاتور استوکاستیک
۴	RSI:اندیکاتور قدرت نسبی
۵	OBV:اندیکاتور حجم متوازن
۶	WILLIAMSR:اندیکاتور آر درصد ویلیامز

تمام درخت‌های مورد استفاده در پژوهش از عمق چهار انتخاب شده‌اند. تعداد درخت‌های جنگل تصادفی برابر ۱۰ درخت انتخاب، و برای فرآیند جداسازی در گره‌ها از آنتروپی استفاده شد. به عنوان نمونه یکی از درخت‌های تصمیم و خروجی آن در شکل ۱ ارایه شده است:



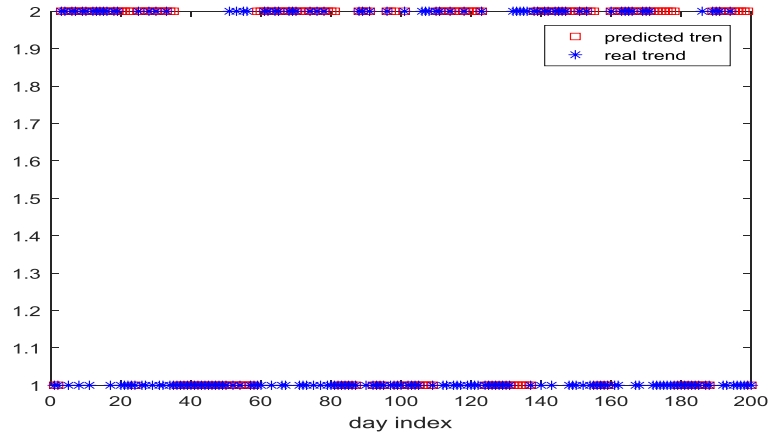
شکل ۱- نمونه درخت تصمیم مورد استفاده در پژوهش

که دارای مشخصات جدول ۲ می‌باشد:

جدول ۲- خروجی یک نمونه از درخت های تصمیم

شماره گره	فیچر مورد استقاده	نقطه برش	خروجی گره
1	5	6.6577E+10	[]
2	5	1.9844E+10	[]
3	2	-1204.6824	[]
4	4	8.99004438	[]
5	4	30.1905728	[]
6	4	[]	1
7	4	26.3330442	[]
8	5	1.7677E+10	2
9	5	1.9844E+10	1
10	4	[]	1
11	5	1313273967	1
12	[]	[]	[]
13	[]	[]	[]
14	4	21.1193667	1
15	5	1313273967	1

بعد از طراحی جنگل تصادفی ،۲۰۰ داده تست وارد جنگل شده و خروجی جنگل تصادفی با خروجی واقعی مقایسه گردید. دقت پیش‌بینی در اجرای اول برابر ۰/۶۰ و در اجرای دوم برابر ۰/۶۴ می باشد. میانگین نهایی جنگل تصادفی برابر با ۰/۶۴ درصد تخمین زده شده است. که از میانگین ۵۰ بار اجرای نرم افزار بدست آمده. علت گرفتن میانگین، استفاده از تکنیک آنتروپی و فیچرهای تصادفی می‌باشد. که به علت تصادفی بودن ممکن است در هر اجرا دقیقا همان جواب اجرای قبل را نداشته باشد. مقدار میانگین نهایی (۰/۶۴) نشان می‌دهد که ۰/۶۴ پیش بینی با مقدار واقعی مطابقت داشته و درست پیش‌بینی شده است. به عنوان نمونه اجرای دوم نرم افزار بر روی داده های تست در شکل ۲ نمایش داده شده. در نمودار مذکور ستاره‌های آبی مقدار واقعی سهام، و مربع‌های قرمز مقدار پیش‌بینی شده‌ی سهام است.



شکل ۲- نمودار حاصل از پیش بینی جنگل تصادفی بر روی دادهای تست در مقایسه با خروجی واقعی.اجرای دوم

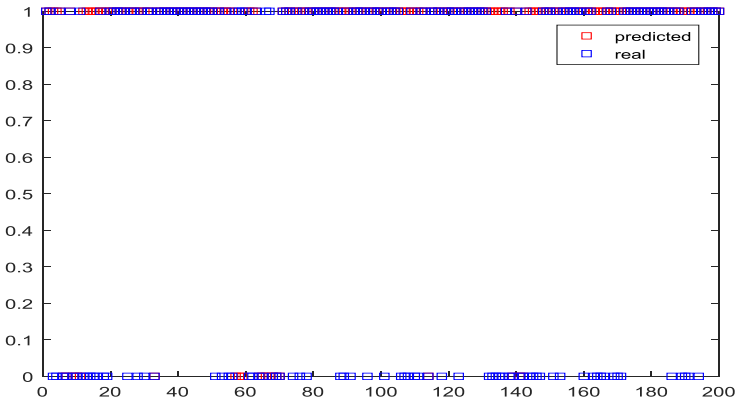
- **تحلیل داده‌ها با روش رگرسیون لجستیک**

در ادامه برای تست و ارزیابی دقت الگوریتم جنگل تصادفی، با استفاده از رگرسیون لجستیک به پیش‌بینی جهت روند بازار پرداخته می‌شود. برای این منظور ابتدا داده‌های آموزش را جهت مشخص شدن ضرایب رگرسیون استفاده نموده، و در ادامه با استفاده از داده‌های تست دقت این روش بررسی می‌شود. نتیجه حاصل از تقریب رگرسیون لجستیک به کمک نرم افزار متلب در جدول ۳ ارایه شده است.

جدول ۳- تقریب رگرسیون لجستیک برای متغیر وابسته جهت بازار

متغیر	مقدار	آماره z	مقدار احتمال
C	0	1.12542	0.0012254
MACD	-0.00029	-2.35554	0.018495624
OBV	5.53E-12	2.519798	0.011742218
PROC	0	0.012853	0.0014524
RSI	-0.00781	-1.49177	0.135760026
SO	0.016497	4.047933	5.17E-05
WILLIAMSR	0.004534	2.243688	0.024852505

مقدار احتمال نشان می‌دهد که در سطح اطمینان ۰,۹۵ تمامی ضرایب به غیر از RSI معنادار می‌باشند. حال به کمک رگرسیون لجستیک به پیش‌بینی جهت بازار برای روز آتی بر روی داده‌های تست اقدام شد. همانگونه که بیان شد خروجی رگرسیون لجستیک به صورت باینری بوده، به این صورت که اگر مقدار رگرسیون کمتر از عدد ۰/۵ باشد، مقدار عددی رگرسیون برابر با صفر و در غیر این صورت برابر یک خواهد بود. که عدد صفر نشان دهنده‌ی پیش‌بینی نزولی بودن بازار در یک روز آینده، و عدد یک نشان دهنده‌ی روند صعودی بودن بازار در یک روز آینده می‌باشد. نتیجه حاصل از پیش‌بینی در شکل ۳ ارایه شده است. که نشان می‌دهد میانگین مقدار تطابق داده‌های پیش‌بینی شده با مقدار واقعی برابر با ۰/۶۲ می‌باشد.



شکل ۳- مقایسه نتایج حاصل از رگرسیون لجستیک بر روی داده های تست با مقدار واقعی

- **تحلیل داده‌ها با روش کاملا تصادفی**

در روش کاملا تصادفی ابتدا دستور اجرا برای ۲۰۰ داده‌ی تست به نرم‌افزار داده شده. به گونه-ای که اگر شیر آمد عدد یک و اگر خط آمد عدد صفر، به داده‌ها تعلق بگیرد. پس از ۵۰ بار اجرا از نرم‌افزار و میانگین‌گیری، خروجی برابر با ۰/۴۹۸۵ محاسبه شده. که نشان می دهد تقریبا نیمی از پیش‌بینی‌ها با مقدار واقعی تطابق داشته است.

۵- نتیجه‌گیری و بحث

همانگونه که اشاره گردید در پژوهش حاضر به پیش‌بینی قیمت سهام با رویکرد جنگل تصادفی پرداخته شده است و با تکیه بر ادبیات و پیشینه تحقیق، ۶ شاخص تکنیکال تاثیرگذار در قیمت سهام را به عنوان ویژگی در جنگل تصادفی در نظر گرفته می‌شود. داده‌های شاخص کل که جامعه-ی آماری پژوهش هستند؛ به دو دسته‌ی داده‌های آموزش و داده‌های تست تقسیم بندی می‌شوند. با استفاده از داده‌های آموزش درختان جنگل تصادفی آموزش داده می‌شوند و در ادامه داده‌های تست وارد درختان آموزش دیده، شده و میزان دقت الگوریتم اندازه‌گیری می‌شود. جهت خوشه-بندی از تکنیک آنتروپی بهره گرفته شده به صورتی که ویژگی‌ای که کمترین میزان بی‌نظمی را ایجاد کند به عنوان نقطه‌ی برش درنظر گرفته می‌شود. و خروجی الگوریتم به صورت باینری می-باشد. در یکبار اجرا از الگوریتم، خروجی ۱۰ درخت تصمیم‌گیری محاسبه می‌شود. در ادامه میانگین نهایی جنگل تصادفی برابر با ۰/۶۴ درصد تخمین زده شده است. که از میانگین ۵۰ بار اجرای نرم‌افزار بدست آمده. علت میانگین گرفتن، استفاده از تکنیک آنتروپی و فیچرهای تصادفی می‌باشد. که به علت تصادفی بودن ممکن است در هر اجرا دقیقا همان جواب اجرای قبل را نداشته باشد. این مقدار نشان می‌دهد که ۰/۶۴ پیش‌بینی قیمت سهام بازار با مقدار واقعی مطابقت داشته و درست پیش‌بینی شده است.

در ادامه و برای مقایسه‌ی دقت الگوریتم با روش‌های خطی و کاملا تصادفی، الگوریتم جنگل تصادفی با الگوریتم رگرسیون لجستیک و روش پرتاب سکه (کاملا تصادفی) مقایسه گردیده است. که دقت هر یک از روش‌ها با میانگین‌گیری از ۵۰ بار اجرا، به ترتیب ۰/۶۲ و ۰/۴۹۸۵اندازه‌گیری شد که نشان دهنده‌ی میزان تطابق مقدار پیش‌بینی شده سهام با مقدار واقعی آن می‌باشد. جدول ۴ نتایج حاصل از سه روش یاد شده، با بهره‌گیری از نرم‌افزار مطلب را نشان می‌دهد.

جدول۴- مقایسه نتایج حاصل از سه روش کاملا تصادفی، روش جنگل تصادفی و رگرسیون لجستیک

	میانگین	انحراف معیار
روش تصادفی	۰,۴۹۸۵	۰,۰۳۵۹
روش جنگل تصادفی	۰,۶۴۲۰	۰,۰۲۵۰
رگرسیون لجستیک	۰,۶۲	-

پس از بررسی نتایج حاصل از سه روش یاد شده به این نتیجه می‌رسیم که الگوریتم جنگل تصادفی در پیش‌بینی سهام برای یک روز آینده از کارایی بالایی برخوردار است و به دلیل سادگی در تفسیر و به کارگیری این الگوریتم، تمامی افراد فعال در بورس تهران و جهان می‌توانند از این روش بهره جویند. همچنین نشان می‌دهد که در تکنیک‌های داده‌کاوی، روش‌های هوش مصنوعی(الگوریتم جنگل تصادفی) نسبت به روش‌های آماری(رگرسیون لجستیک) از دقت بالاتری برخوردار بوده و روش‌های غیرخطی نسبت به روش‌های خطی برتری دارد. دیگر تحقیقات صورت گرفته نیز، صحت این ادعا را تایید می‌کنند.

خایدم و همکاران، (۲۰۱۶) در تحقیقی تحت عنوان "پیش‌بینی روند قیمت در بازار سهام با استفاده از الگوریتم جنگل تصادفی، سهام سه شرکت معتبر از جمله سامسونگ، اپل و جنرال الکتریک را پیش‌بینی کرده. سهام هر شرکت برای یک ماه آینده پیش‌بینی شده است. داده‌های سهام نیز با تکنیک هموار سازی نمایی هموار شده. و داده‌های پرت خذف شده و دقتی با میانگین حدود ۸۷ درصد را به خود اختصاص داده است. که در برابر دقت ۶۴/۲ درصدی پژوهش حاضر از دقت بالاتری برخوردار است. که علت این بالا بودن دقت ۱- دوره‌ی زمانی پیش‌بینی، که قیمت سهام را برای یک ماه آینده پیش‌بینی کرده این درصورتی است که تحقیق حاضر سهام را برای یک روز آینده پیش‌بینی کرده و با توجه به مدت زمان پیش‌بینی دقت بسیار بالایی است. ۲- داده‌های مقاله با استفاده از تکنیک هموارسازی نمایی، هموار شده یعنی داده‌های پرت حذف شده که این خود باعث بالا رفتن دقت پیش‌بینی می‌شود.

همچنین نتایج تحقیقات دیگر نشان می‌دهد که روش رگرسیون لجستیک که توسط لی و یانگ[۱۹] (۲۰۱۴) مطرح شد، دقتی برابر با ۵۵/۶ را دارا بود. دیا و ژانگ (۲۰۱۳) با استفاده از مدل رگرسیون بردار پشتیبان (svm)، رگرسیون لجستیک و تجزیه و تحلیلی درجه دوم برای پیش‌بینی نتایج برای n روز آینده به دقتی برابر با ۴۴/۵۲ تا ۵۸/۳ درصدی دست یافتند؛ و بالاترین دقتی که برای پیش‌بینی قیمت سهام برای یک ماه آینده گزارش شد، روش svm بود که توسط دیا و ژینجی

(۲۰۱۴) ارائه گردید، که مدل دقتی برابر با ۷۹/۳ درصد را دارا بود. با توجه به پیشینه‌ی تحقیق مدل جنگل تصادفی با دقتی معادل با ۸۷ درصد برای پیش‌بینی قیمت در یک ماه آینده از دقت عمل بالاتری در تکنیک‌های خطی (رگرسیون لجستیک) و تکنیک‌های هوش مصنوعی (شبکه های عصبی ،svm ..)برخوردار است. یو و ونجوان (۲۰۱۰) در تحقیق خود از روش C5.0 که یکی از روش های درخت تصمیم‌گیری است استفاده کردند که دقت۹۵ درصد را به خود اختصاص داد. این تحقیق بالا بودن کارایی درخت تصمیم‌گیری در انجام محاسبات را بررسی می‌کند. نوری جلیانی(۱۳۹۱)در تحقیقی تحت عنوان" استفاده از جنگل تصادفی در پیش‌بینی شاخص کل بازار بورس تهران" که در این تحقیق با بکارگیری متغییرهایی از جمله حجم مبنی و نسبت قیمت سهام به دارایی‌های کل بعنوان ویژگی‌ها در درخت تصمیم‌گیری به پیش بینی بازار بورس پرداخته و دقت مدل ۶۲ درصد تخمین زده شده است. همچنین این مدل با روش شبکه‌ی عصبی مقایسه شده که نتایج بدست آمده نشان دهنده‌ی برتری مدل جنگل تصادفی دارد. این پژوهش بدلیل استفاده از اندیکاتورهای متفاوت در پیش‌بینی قیمت سهام با تحقیق حاضر تطابق ندارد اما برتری روش جنگل تصادفی نسبت به روش‌های مشابه داده‌کاوی مانند شبکه‌های عصبی را نشان می‌دهد.

با توجه به مطالب فوق پیشنهادات اجرایی زیر ارائه می‌گردد:

۱) تحقیق حاضر با توجه به نتایج حاصله و برتری نسبت به روش‌های رگرسیون لجستیک و شبکه‌های عصبی و روش کاملا تصادفی دارای اطلاعات مفیدی می‌باشد و از دقت بالایی برخوردار است، از این رو توصیه می‌شود محققین از این روش در پیش‌بینی قیمت سهام استفاده کرده، و متناسب با وابستگی ساختاری که مسئله با عمق درختان، تعداد و نوع فیچرها و روش‌های خوشه‌بندی دارد؛ و همچنین با توجه به اطلاعات تاریخی هر سهم فاکتورهای مورد استفاده را بهینه کرده و در پیش‌بینی به کار ببرند.

۲) این پژوهش به کارشناسان تحلیل تکنیکال در شرکت‌های سرمایه‌گذاری وسایر نهادهای مالی فعال در زمینه سهام و بورس به منظور بهرمندی از قابلیت‌های این روش تحلیل برای کسب سود بیشتر، همچنین به محققین و دانشجویان جهت انجام تحقیقات آتی خود با بهره‌گیری از نتایج پژوهش توصیه می‌شود.

توصیه برای تحقیقات آتی به شرح ذیل ارائه می‌گردد

۱) به محققین توصیه می‌شود، بازه‌ی زمانی داده‌های مورد استفاده را از ۳ سال به ۵ سال یا بیشتر توسعه داده و نتایج را با نتایج تحقیق حاضر مقایسه کنند.

۲) محققین از دیگر اندیکاتورها از جمله شاخص کانال کالا یا(CCI)[۲۰] ، باندهای بولینگر[۲۱]، ADX[۲۲]و غیره در پیش‌بینی قیمت به کمک الگوریتم جنگل تصادفی، استفاده کنند و نتایج بدست آمده را با نتایج تحقیق حاضر مقایسه کنند.

۳) اندیکاتورهای مورد استفاده تحقیق را در روش‌های دیگر داده‌کاوی مانند شبکه‌های عصبی یا رگرسیون بردار پشتیبان به کارگیرند، و نتایج را با پژوهش حاضر مقایسه کنند.

۴) محققین می‌توانند، از دیگر روش‌های خوشه بندی مانند معیار اینفورمیشن جین[۲۳]، معیار جینی ایندکس[۲۴]، جین ریت[۲۵]، معیار DKM استفاده کرده، و نتایج را با پژوهش حاضر مقایسه کنند.

۵) محققیق با بهره گیری از این روش قیمت سهام را برای یک ماه تا یک سال آینده پیش-بینی کنند. یا داده‌های تاریخی سهام را با تکنیک هموارسازی نمایی هموار کرده، و نتایج را با نتایج تحقیق حاضر مقایسه کنند.

۶) محققین در هر درخت تصمیم درصد جمعیت وارده به هر گره را کم یا اضافه کرده، و نتایج را با تحقیق حاضر مقایسه کنند.

فهرست منابع

* آهنگری مهناز. بکارگیری الگوریتم درخت تصمیم جهت پیش‌بینی شرکت‌های ورشکسته و غیر ورشکسته پذیرفته شده در بورس اوراق بهادار تهران. ۱۳۹۰.اولین کنفرانس ملی دانش پژوهان کامپیوتر و فناوری اطلاعات، تبریز، دانشگاه تبریز.
* البرزی محمود. بکارگیری تکنی های خوشه‌بندی و الگوریتم ژنتیک در بهین سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک‌ها. فصلنامه آینده پژوهی مدیریت. ۱۳۹۱. دوره ی ۱. شماره ۱ .
* بدیعی حسین. روح‌الله بدیع‌زاده. هادی محمودی. پیش‌بینی رفتار قیمت در بورس اوراق‌بهادار با استفاده از شبکه‌ی عصبی مصنوعی. مجله‌ی مهندسی مالی و مدیریت اوراق بهادار. شماره ۳۱. ۱۳۹۶.
* تقی‌پوریان یوسف. رئیس‌زاده سید محمدرضا. فلاح عابد آسیه. رابطه بین ارزیابی عملکرد مالی و پیش‌بینی ورشکستگی در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. اولین کنفرانس بین‌المللی حسابداری، مدیریت و نوآوری در کسب و کار، گیلان. ۱۳۹۴.
* حقیقت منفرد جلال. محمود احمدعلی نژاد. سارا متقالچی. مقایسه مدل‌های شبکه عصبی با مدل سری زمانی باکس-جنکینز در پیش‌بینی شاخص کل قیمت سهام بورس اوراق بهادارتهران. مجله‌ی مهندسی مالی و مدیریت بورس اوراق بهادار. شماره‌ی یازدهم. تابستان ۱۳۹۱.
* زراء نژاد منصور. حمید شهرام. پیش‌بینی نرخ تورم در اقتصاد ایران با استفاده از شبکه‌های عصبی مصنوعی پویا(دیدگاه سری زمانی .). فصلنامه اقتصاد مقداری. دانشگاه شهید چمران اهواز.۱۳۸۸. شماره ۶. صص۱۶۷-۱۴.
* سعادت محمد. زمانی بروجنی فرساد. مروری بر روش های بهبود کارایی تکنیک جنگل تصادفی. اولین همایش ملی فناوری اطلاعات، ارتباطات و محاسبات نرم اصفهان دانشگاه آزاد اسلامی واحد اصفهان (خوراسگان).۱۳۹۵.
* شهریاری شهروز. بازرگان عبدالرضا. عباسی نرگس. یادگیری درخت تصمیم‌گیری. دومین همایش منطقه ای علوم ریاضی و کاربردها، تنکابن، دانشگاه پیام نور استان مازندران.۱۳۹۳
* علیزاده سمیه. داده کاوی وکشف دانش گام به گام با نرم‌افزار، انتشارات دانشگاه صنعتی خواجه نصیرالدین طوسی.۱۳۹۳.
* فروغی داود. زیبانژاد یاسر. مونادجمعی اکبر. اجرای فن درخت تصمیم برای پیش‌بینی ورشکستگی. دانش کامپیوتر و مهندسی اتوماسیون.۱۳۹۰ ۴. صص ۱۶۵-۱۶۹.

* کامروافر محمد.سید ذبیح‌الله هاشمی. بررسی و شناخت متغیرهای اصلی تاثیرگذار بر شاخص کل بورس اوراق‌بهادار تهران و مدلسازی آن با استفاده از شبکه‌های عصبی مصنوعی و مقایسه نتایج حاصله با تحلیل تکنیکال و موجهای الیوت. مجله‌ی مهندسی مالی و مدیریت اوراق‌بهادار. شماره سی‌ام. بهار ۱۳۹۶.
* مهرآرا اسداله. زهرا عطف و زهرا عسگری. تکنیک‌های داده‌کاوی در پیش‌بینی قیمت سهام بورس اوراق‌بهادار، کنفرانس ملی حسابداری. مدیریت مالی و سرمایه‌گذاری، گرگان، دانشگاه جامع علمی کاربردی استان گلستان.۱۳۹۱.
* نوری جلیلانی محمد حسن. استفاده از روش جنگل تصادفی در پیش‌بینی شاخص کل بورس ایران. پایان‌نامه کارشناسی ارشد «علوم تصمیم ومدیریت دانش» دانشگاه خوارزمی. دانشکده فنی مهندسی ص ۱تا ۲۰. ۱۳۹۱.
* محمدی .شاپور. تحلیل تکنیکی در بورس اوراق بهادار تهران. مجله‌ی تحقیقات مالی .سال ششم.۱۳۸۳. شماره . ۱۷.
* Adebiyi ,Ayodele .Stock Price Prediction using Neural Network with Hybridized Market Indicators. Journal of Emerging Trends in Computing and Information Sciences.(2011). VOL. 3, NO. 1,pp.1-9
* Breiman, L. Friedman, J. H., Olshen. R. A., & Stone, C. J. Classification and regression trees. New York: Chapman & Hall/CRC.1984
* Finger, C. The ability of earnings to predict future earnings and cash flows. Journal of Accounting Research.1994.., 32: 210-223.
* Haoming Li, Zhijun Yang and Tianlun Li . Algorithmic Trading Strategy Based On Massive Data Mining. Stanford University. 2014
* Leo Breiman .Statistics Department ,University of California Berkeley, CA 94720. Random Forests. 2001
* Wang, H., Jiang, Y., & Wang, H. .Stock return prediction based on Bagging-decision tree. In 2009 IEEE International Conference on Grey Systems and Intelligent Services: 1575-1580. IEEE.
* Xinjie. Stock Trend Prediction With Technical Indicators using SVM. Stanford University. 2014
* Yu, G., & Wenjuan, G.. Decision tree method in financial analysis of listed logistics companies. International conference on intelligent computation technology and automation. 2010
* Yuqing Dai, Yuning Zhang. Machine Learning in Stock Price Trend Forecasting. Stanford University. 2013
* Khaidem L . Saha S. Roy Dey S. Predicting the direction of stock market prices using random forest. To appear in Applied Mathematical Finance Vol. 00, No. 00, Month 20XX, 1-20 .2016

یادداشت‌ها

[1] Khaidem
[2] ARMA
[3] Breiman
[4] Overfitting
[5] Bias
[6] Bagging
[7] Ensemble
[8] Xinjie
[9] Dai & Zhang
[10] Adebiyi ,Ayodele
[11] Yu & Wenjuan
[12] Wang
[13] MATLAB
[14] Moving Average Convergence Divergence
[15] Relative Strengh Index
[16] Stochastic Oscilliator
[17] Williams %R
[18] On Balanced Volume
[19] Li & Yang
[20] Commodity Channel Index
[21] Bollinger Bands
[22] Average Directional Movement Index
[23] Information Gain
[24] Gini Index
[25] Gain Ratio