فصلنامه پژوهش‌های پولی_ بانکی
سال سیزدهم، شماره ۴۵، پاییز ۱۳۹۹
صفحات ۴۳۵_ ۴۶۰

رتبه‌بندی سهام شرکت‌های بورس اوراق بهادار تهران بر اساس مدل ترکیبی درخت تصمیم و رگرسیون لجستیک

رضا تهرانی [*] **زهرا نیکخواه بهرامی** [†]

تاریخ دریافت: ۱۳۹۸/۰۸/۱۴ تاریخ پذیرش: ۱۳۹۹/۰۹/۱۶

چکیده

تاکنون تحقیقات بسیاری در چهارچوب مدل‌های خطی یا غیرخطی و با استفاده از مدل‌های آماری و ابزارهای یادگیری ماشین در هوش مصنوعی برای برآورد نرخ بازده سهام در ایران معرفی شده است. هدف عمدهٔ این روش‌ها استفادهٔ هم‌زمان از متغیرهای مستقل متفاوت برای بهبود مدل‌سازی نرخ بازده سهام است؛ درحالی‌که در فرایند پیش‌بینی‌پذیری نرخ بازده، علاوه بر نحوهٔ مدل‌سازی، میزان همبستگی متغیرهای مستقل با یکدیگر و درنتیجه افزایش اریبی برآوردگرهای مدل نیز از اهمیت ویژه‌ای برخوردار است. ازاین‌رو، در این مقاله بر اساس مدل ترکیبی درخت تصمیم و رگرسیون لجستیک به‌صورت هم‌زمان متغیرهای اثرپذیر را تشخیص داده شده و سپس مدلسازی غیرخطی نرخ بازده انجام شده است. به‌منظور بررسی مدل پیشنهادی، اطلاعات ۱۰۰ شرکت پذیرفته‌شده در بورس اوراق بهادار طی بازهٔ زمانی ۱۳۹۷-۱۳۹۰ در نظر گرفته و بر اساس مدل پیشنهادی، وزن‌های انتخاب پرتفوی بهینه برآورد شده است. نتایج بررسی ما نشان می‌دهد که الگوریتم ترکیبی پیشنهادی، از مدل‌های رقیب بازدهی بهتری دارد.

واژه‌های کلیدی: رتبه‌بندی سهام، درخت تصمیم، رگرسیون لجستیک، الگوریتم ترکیبی
طبقه‌بندی JEL: C38, C44, G11

[*] استاد، گروه مدیریت مالی و بیمه، دانشکدهٔ مدیریت، دانشگاه تهران، تهران، ایران؛ rtehrani@ut.ac.ir
[†] دانشجوی دکتری مدیریت مالی، گروه مدیریت مالی و بیمه، دانشکدهٔ مدیریت، دانشگاه تهران، تهران، ایران؛ z.nikkhahbahrami@ut.ac.ir (نویسندهٔ مسئول)

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

۱ مقدمه

انتخاب سهام ازجمله مهم‌ترین و حیاتی‌ترین تصمیمات افراد حقیقی و حقوقی سرمایه‌گذاران در بورس اوراق بهادار تهران است که می‌توان گفت از گذشته تاکنون، یکی از مسائل مهم موردبحث بوده است و نیز با پژوهش‌هایی که در این زمینه صورت گرفته است، مدل‌هایی برای تعیین پرتفوی ارائه شده است که به مرور زمان، ایرادات هرکدام مشخص و مدل‌های دیگر جایگزین آن شده است (لشکری و نظام‌الاسلامی، ۱۳۹۵).

پیش‌بینی قیمت و یا بازده سهام جزء مباحث مالی است که توجه محققان بسیاری را طی سال‌های متمادی به خود جلب کرده است. با توجه به اینکه بازده سهام متأثر از متغیرهای متعدد است، لذا پیش‌بینی آن کار ساده‌ای نیست (داگلی و همکاران، ۲۰۰۳). در واقع، می‌توان گفت یکی از مسائل مهمی که پژوهشگران و دانشمندان حوزهٔ تصمیم‌گیری و پیش‌بینی با آن روبه‌رو هستند، انتخاب متغیرهای تأثیرگذار در خروجی تصمیم‌گیری و پیش‌بینی است؛ بنابراین، اگر بتوان بازده سهام را با استفاده از متغیرهای مناسب پیش‌بینی کرد و مدل‌هایی برای آن ارائه داد، می‌توان شرایط مطمئن‌تری در بازار سرمایه ایجاد کرد که به گسترش سرمایه‌گذاری در بازارهای مالی کمک خواهد کرد (لشکری و نظام‌الاسلامی، ۱۳۹۵).

مدیران و سرمایه‌گذاران حقیقی و حقوقی به‌دلیل وجود انبوه متغیرهای تأثیرگذار ترجیح می‌دهند سازوکاری در اختیار داشته باشند که بتواند آن‌ها را در امور تصمیم‌گیری‌شان یاری و مشاوره دهد؛ به همین دلیل، به روش‌های پیش‌بینی روی می‌آورند که به‌واسطهٔ آن‌ها تخمین‌هایشان به واقعیت نزدیک و خطاهایشان بسیار کم باشد (ایمانی، ۱۳۹۰). یکی از روش‌های یادگیری ماشین که برای انتخاب متغیرهای تأثیرگذار قابل‌استفاده است، روش‌های مبتنی بر درخت تصمیم است، زیرا داده‌های بازار مالی عموماً دارای خطاها و نوسانات بسیاری است و مدل‌های خطی به‌دلیل آثار متقابل زیادی که بین متغیرها وجود دارد، توانایی مدل‌سازی بهینه آن‌ها را ندارند. همچنین نبود وابستگی متغیرهای مستقل از دیگر فرضیاتی است که عموماً در مدل‌های خطی پارامتری سری‌های زمانی فرض می‌شود. درحالی‌که در داده‌های واقعی متغیرهای پیشگوی نرخ بازده وابستگی زیادی با یکدیگر دارند که به‌سادگی نمی‌توان از آن‌ها در مدل‌سازی نرخ بازده استفاده کرد.

به‌طور خلاصه نوآوری این پژوهش، استفاده از مدل غیرخطی ترکیبی است که بر اساس انعطاف‌پذیری آن، هم امکان مدل‌بندی متغیرهایی با اثر غیرخطی و هم شناسایی محرک‌های اصلی در پیش‌بینی بازده سهام را دارد. زیرا یکی از چالش‌های افراد باتجربه در تحلیل نرخ

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

بازدهی مقطعی، همواره شناسایی محرک اصلی نرخ بازدهی سهام در بین تعداد زیادی متغیر بوده است. لذا ضروری است که از روش‌هایی به‌جز سبک رگرسیون فاما_فرنچ نیز استفاده کرد (کاکران، ۲۰۱۱)؛ بنابراین، مدل ارزیابی نظام‌مند بر مبنای درخت تصمیم برای پیش‌بینی نرخ بازده سهام مقطعی در مدل پیشنهادی ارائه شده است که در مقایسه با تکنیک‌های قبلی موجود، انتخاب بهینهٔ متغیرهای مستقل به‌صورت هم‌زمان با مدل‌بندی انجام نمی‌شود.

در این پژوهش به‌منظور انتخاب بهترین متغیرهای پیش‌بینی‌کننده نرخ بازده سهام مدل پیشنهادی با ترکیب روش درخت‌های رگرسیونی و رده‌بندی (کارت)[1] (برای رده‌بندی مشاهده‌ها) و رگرسیون لجستیک (برای رده‌بندی متغیرها بر اساس میزان اثر آن‌ها در نرخ بازده) ارائه می‌شود. متغیر اصلی که مدل‌سازی برای آن انجام می‌شود، تفاوت بازده شرکت‌های با بازدهی مثبت با شرکت‌های با بازدهی منفی است. کاربرد این روش ترکیبی در مقایسه با هریک از روش‌های کارت و رگرسیون لجستیک هم در داده‌های شبیه‌سازی‌شده و هم در داده‌های بازده سهام شرکت‌های موجود در بورس اوراق بهادار تهران، موردارزیابی قرار خواهد گرفت.

۲ ادبیات موضوع

اگرچه تکنیک‌های غیرخطی درحال‌حاضر به‌طور گسترده در بین محققین حوزه مالی استفاده نمی‌شود، اما این تکنیک‌ها برای متنوع‌سازی سهام جذاب به نظر می‌رسد (عسکری‌راد، ۱۳۹۰). بااین‌وجود، روش کارت در بازارهای مالی برای رده‌بندی شرکت‌های درمانده‌شده از لحاظ مالی (فریدمن[2] و همکاران، ۱۹۸۵)، تخصیص دارایی (سورنسن و همکاران[3]، ۱۹۹۸)، و انتخاب سهم (سورنسن و همکاران، ۲۰۰۰) استفاده شده است. به‌علاوه تعداد زیادی از تکنیک‌های غیرخطی برای مسئلهٔ پیش‌بینی نرخ بازده استفاده شده‌اند یا اینکه قابلیت بالقوه را در این موضوع دارند.

استفاده از مدل‌های غیرپارامتری و غیرخطی از محبوبیت بیشتری در این حوزه برخوردار است. به‌عنوان مثال در عمل، تکنیک‌هایی که توابع کلاس‌بندی[4] بر اساس درخت تصمیم

---

[1] Classification And Regression Trees (CART)

[2] Friedman

[3] Sorensen et al.

[4] Classification

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

به‌منظور رده‌بندی متغیرها تعریف می‌کنند، از عملکرد بهتری برخوردارند (بریمن و همکاران[1]، ۱۹۸۴). یکی از تکنیک‌های مبتنی بر درخت تصمیم‌گیری، روش درخت‌های رگرسیونی و رده‌بندی (کارت) است که اولین بار توسط بریمن و همکاران (۱۹۸۴) پیشنهاد شد. این روش در واقع تکنیکی غیرخطی و غیرپارامتری است و فرضیه‌های دشواری را که در روش‌های رگرسیون کلاسیک وجود دارد (مانند نرمال بودن داده‌ها)، به مدل‌سازی تحمیل نمی‌کند. روش کارت روشی استوار، انعطاف‌پذیر، و توزیع آزاد است؛ اما هنوز این روش به‌طور گسترده در جامعهٔ سرمایه‌گذاران مورداستفاده قرار نگرفته است. ذکر این نکته ضروری است، مدل‌های معمولی درخت تصمیم برای داده‌های سری زمانی کاربردی ندارد و باید از درخت‌های تصمیم که بر اساس داده‌های پانلی طراحی شده است، استفاده شود. لذا در این مقاله از مدل درخت تصمیم پانلی (طولی) به‌منظور مدل کردن متغیر زمان در درخت تصمیم استفاده کرده‌ایم.

ازآنجاکه یکی از چالش‌های پیش‌ِروی محققین حوزه مالی در تحلیل نرخ بازدهی مقطعی، شناسایی محرک اصلی نرخ بازدهی سهام در بین تعداد زیادی متغیر است (کاکران[2]، ۲۰۱۱)؛ لذا این مدل ارزیابی نظام‌مند می‌تواند به‌صورت درختی برای پیش‌بینی نرخ بازده سهام مقطعی مورداستفاده قرارگیرد. با توجه به اینکه روش کارت به‌طور مستقیم کارایی لازم را در انتخاب متغیرهای مستقل در داده‌های حوزهٔ مالی و بورس اوراق بهادار ندارد، لذا در این مقاله روش هیبریدی جدیدی را که توسط ژو[3] و همکاران (۲۰۱۲) معرفی شده است، ارائه خواهیم داد و سپس با استفاده از تکنیک‌های غیرخطی توزیع آزاد که فرض توزیعی بر روی داده‌ها در نظر نمی‌گیرند، مدل‌بندی بازده‌های سهام را انجام می‌دهیم.

۳ پیشینهٔ پژوهش

در سال‌های اخیر، استفاده از شبکه‌های عصبی مصنوعی و یادگیری عمیق در پیش‌بینی بازده سهام در بازارهای مالی جهان، به‌طور فزاینده‌ای افزایش یافته است. در ایران نیز محققان و فعالان بازار به استفاده از این روش‌های نوین در پیش‌بینی بازار سرمایه پرداخته‌اند. در پژوهش راعی و چاوشی (۱۳۸۲) به‌منظور پیش‌بینی بازده سهام در بورس اوراق بهادار تهران با استفاده از مدل شبکه‌های عصبی مصنوعی و مدل چندعاملی نشان داده شد که این دو مدل در پیش‌بینی رفتار بازده سهام موفق است؛ اما عملکرد شبکه‌های عصبی مصنوعی بر مدل

[1] Brieman et al.
[2] Cochrane ،J.
[3] Zhu, M. D

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

چندعاملی برتری دارد. در تحقیقات نمازی و کیامهر (۱۳۸۶) میزان کارایی شبکهٔ عصبی در پیش‌بینی نرخ بازده موردبررسی قرار گرفت و نتایج آن‌ها نشان داد، شبکهٔ عصبی مصنوعی توانایی پیش‌بینی بازده روزانهٔ سهام را با خطایی نسبتاً مناسب دارد. قلی‌زاده و وحیدپور (۱۳۸۶) با استفاده از روش خودرگرسیون همبستهٔ برداری، برآورد و پیش‌بینی نهایی مدل رگرسیونی پیش‌بینی بازده را بهبود دادند. قالیباف اصل و معصوم‌زاده (۱۳۸۸) به‌منظور بررسی پیش‌بینی احتمال تغییر قیمت سهام با استفاده از رگرسیون لجستیک در بورس اوراق بهادار تهران نشان دادند که میان قیمت‌های گذشتهٔ سهام، درصد سهام مبادله‌شده، شاخص بازده نقدی، و قیمت و تغییرات آتی قیمت سهام رابطه‌ای معنی‌دار وجود دارد و ثانیاً رابطه میان قیمت‌های روز گذشته سهام و تغییرات روز معاملاتی بعد بسیار قوی بوده است و تقریباً در تمام شرکت‌های موردبررسی مصداق دارد.

محمدی و سعیدی (۱۳۹۱) به‌منظور پیش‌بینی نوسانات بازده بازار با استفاده از مدل ترکیبی گارچ و شبکهٔ عصبی این نتیجه رسیدند که سری زمانی داده‌های بازده دارای سه ویژگی نوسان خوشه‌ای، عدم تقارن، و غیرخطی است. نتایج آن‌ها نشان داد که مدل‌های ترکیبی همسویی بیشتری با نوسان واقعی نسبت‌به مدل‌های پایه‌ای گارچ دارد. راعی و آرا (۱۳۹۱) با مقایسهٔ کارایی مدل‌های رگرسیون با رویکرد تحلیل مؤلفه‌های اصلی[1] و شبکه‌های عصبی مصنوعی در پیش‌بینی بازده غیرعادی نشان دادند که توانایی مدل شبکه‌های عصبی مصنوعی پیش‌خور با الگوریتم پس‌انتشار خطا در پیش‌بینی برون نمونه‌ای بازده غیرعادی سهام موردمعامله در بورس اوراق بهادار تهران در سال‌های ۱۳۸۴ تا ۱۳۹۱ به‌طوری معنادار بیشتر از توانایی رگرسیون خطی با رویکرد تحلیل مؤلفه‌های اصلی بوده است. فشاری و مظاهری‌فر (۱۳۹۵) به مقایسهٔ دو روش شبکهٔ عصبی مصنوعی و شبکهٔ فازی‌ـ‌عصبی در حل مسئلهٔ بهینه‌سازی پرتفوی پرداخته و نشان داده‌اند که الگوریتم شبکهٔ عصبی می‌تواند روشی قابل‌اتکا برای سهام‌داران باشد. ایدهٔ اصلی این مقاله بهره‌گیری از مدلی است که بتواند اولاً مراحل آن قابل‌توصیف باشد و ثانیاً آثار سری زمانی متغیرها در مدل را در نظر بگیرد. در اغلب روش‌های شبکه‌های عصبی که تاکنون در مقالات کشور برای داده‌های بورس اوراق بهادار پیشنهاد شده است، ویژگی زمانی متغیرها مدل‌سازی نشده است.

دش[2] و همکاران (۲۰۱۹) در پژوهشی به طبقه‌بندی بر اساس مدل تاپسیس یکپارچه، به پیش‌بینی حرکت قیمت سهام شاخص پرداختند. در این پژوهش، مجموعهٔ طبقه‌بندی‌شدهٔ

[1] Principal component analysis, PCA
[2] Dash

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

رتبه‌بندی مبتنی بر مدل تاپسیس برای پیش‌بینی شاخص قیمت سهام ارائه شده است. تکنیک رتبه‌بندی ترجیح بر اساس شباهت به راه‌حل تاپسیس، یکی از تکنیک‌های محبوب تصمیم‌گیری چندمعیاره، برای رتبه‌بندی و انتخاب مجموعه‌ای از طبقه‌بندی‌کننده‌های پایه برای گروه پیشنهاد شده است. درحالی‌که وزن طبقه‌بندی‌کننده‌های مورداستفاده در گروه با استفاده از روش تاپسیس تنظیم می‌شود، مدل گروه پیشنهادی برای پیش‌بینی شاخص قیمت سهام نسبت به قیمت‌های قبلی شاخص‌های سهام BSE SENSEX، S & P500، و NIFTY 50 اعتبار دارد. این مدل در مقایسه با طبقه‌بندی‌های فردی و سایر مدل‌های گروه مانند اکثریت رأی، اکثریت وزنی، دیفرانسیل تکاملی، و گروه طبقه‌بندی مبتنی بر بهینه‌سازی عملکرد بهتری نشان داده است.

ژو[1] و همکاران (۲۰۱۷) در تحقیقی به پیش‌بینی وضعیت لیست شرکت‌های دارای لیست چینی با استفاده از درخت تصمیم‌گیری همراه با روش انتخاب ویژگی فیلتر بهبودیافته پرداختند. آن‌ها روش انتخاب ویژگی فیلتر بهبودیافته را پیشنهاد کردند تا ویژگی‌های مؤثری را برای پیش‌بینی وضعیت لیست شرکت‌های دارای فهرست چینی انتخاب کند. با توجه به نگرانی‌های عملی تحلیلگران در امور مالی در مورد عملکرد و تفسیر مدل‌های پیش‌بینی، مدل‌های مبتنی بر درخت تصمیم‌گیری C4.5 و C5.0 استفاده شده است. برای ارزیابی استحکام مدل‌ها با زمان، مدل‌ها نیز در دورهٔ زمانی مشخص موردآزمایش قرار می‌گیرد. نتایج تجربی اثربخشی روش انتخاب ویژگی پیشنهادی و مدل درخت تصمیم C5.0 را نشان می‌دهد. با توجه به پیشینهٔ پژوهش‌های انجام‌شده می‌توان گفت، مزیت مدل پیشنهادی این مقاله این است که با بهره‌گیری از رگرسیون مقطعی که برای داده‌های پانلی مورداستفاده قرار می‌گیرد، در همهٔ مراحل درخت تصمیم سری زمانی داده‌ها مدل‌سازی خواهد شد. لذا فرضیهٔ این مقاله این است که:

- مدل ترکیبی درخت تصمیم عملکرد بهتری نسبت به روش‌های کارت و رگرسیون لجستیک در رتبه‌بندی سهام شرکت‌های بورس اوراق بهادار تهران دارد.

در ادامهٔ مقاله، ابتدا توضیحات مختصری در خصوص مدل‌های درختی و رگرسیون لجستیک ارائه شده است و سپس ایدهٔ اصلی مقاله در معرفی مدل ترکیبی جدید از این دو روش ارائه شده است.

[1] Zhou

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

۴ روش‌شناسی پژوهش

این پژوهش در زمرهٔ پژوهش‌های همبستگی قرار دارد و از طرفی دیگر، پژوهش حاضر از نوع پس‌رویدادی است، یعنی بر مبنای تجزیه‌وتحلیل اطلاعات گذشته (صورت‌های مالی شرکت‌ها) انجام می‌گیرد. جامعهٔ آماری این پژوهش کلیهٔ شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران است که به‌منظور بررسی مدل پیشنهادی در این مقاله، ۱۰۰ شرکت به‌عنوان نمونهٔ آماری موردمطالعه قرار گرفته که برای انتخاب نمونه از روش غربالگری (حذفی) استفاده شده و برای این منظور معیارهای زیر در نظرگرفته شده است: ۱ـ حداقل ۲ سال قبل از سال ۱۳۹۰ سهام آن‌ها در بورس اوراق بهادار عرضه شده باشد؛ ۲ـ شرکت‌های نمونه نباید در خلال دورهٔ زمانی موردمطالعه، از بورس خارج شده باشند؛ ۳ـ سال مالی آن‌ها مطابق با ۲۹ اسفندماه باشد؛ و ۴ـ حذف شرکت‌های لیزینگ، بانک، هلدینگ، و بیمه از نمونه به‌دلیل نوسانات خاص‌تر و غیرقابل‌پیش‌بینی بودن آن‌ها. لذا قابل‌ذکر است، آزمون مدل در این تحقیق، بر اساس سهام شرکت‌هایی است که نتایج مدل را به‌دلیل ناهمگنی تغییرات، تحت تأثیر قرار ندهند. همچنین ویژگی‌های فصلی آن‌ها در سال‌های ۱۳۹۰ الی ۱۳۹۷ از سامانه‌های اطلاعاتی بورس اوراق بهادار تهران که در مجموع ۲۸۰۰ سطر اطلاعاتی است، استخراج شده است. سپس عملیاتی را که در ادامه به‌صورت جزئی‌تر توضیح داده‌شده، بر روی داده‌های استخراج‌شده پیاده‌سازی شده است.

۵ درخت تصمیم کارت و رگرسیون لجستیک

روش کارت یکی از روش‌هایی است که به‌صورت مکرر داده‌ها را به بخش‌های کوچک‌تر تبدیل می‌کند. اولین و مهم‌ترین گام در تعیین گره‌های درخت تصمیم تعریف ریشهٔ آن است که باید در ابتدا در مورد آن تصمیم گرفت. بعد از تعیین ریشهٔ درخت، تمام قوانین برای تولید شاخه‌های بعدی درخت بر اساس قواعد شرطی اتفاق می‌افتد. در واقع، پس از تولید شاخه‌های این درخت می‌توان به‌راحتی رده‌بندی تولیدشده برای هر سهم جدید را استخراج کرد. به کمک این روش به‌راحتی می‌توان حجمی عظیم از داده را در ساختارگرافیکی درختی به نمایش گذاشت که هرکس بتواند از آن به‌راحتی استفاده کند. درحالی‌که مدل‌بندی رگرسیون لجستیک یا دیگر روش‌های غیرخطی برای رده‌بندی سهام نمی‌تواند به‌راحتی توسط فرد غیرمتخصص درک و استفاده شود. همچنین ازآنجایی‌که این روش از مدل‌های غیرپارامتریک محسوب می‌شود، فرضیه‌ای خاص را بر روی داده‌ها فرض نمی‌کند و لذا محدودیت استفاده از آن در مقایسه با روش‌های دیگر بسیار کمتر است.

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

این روش در مقایسه با دیگر روش‌های رگرسیونی و تحلیل ممیزی، پایداری بیشتری در مقابله با داده‌های پرت و خطاهای فراوان دارد. اما یکی از معایب این روش این است که به‌دلیل گسسته در نظر گرفتن متغیرها در کارت حساسیت بیشتری بر روی متغیرهایی که با مقیاس پیوسته اندازه‌گیری شده است، ایجاد می‌کند (ژو و همکاران[1]، ۲۰۱۱). به‌عنوان مثال با اندکی تغییر بر روی یک متغیر پیوسته که بر روی ۱۰ شاخهٔ درخت تقسیم‌بندی شده است، تغییرات بزرگ در رده‌بندی دیده می‌شود. یکی دیگر از معایب استفاده از درخت‌های تصمیم‌گیری این است که به‌دلیل فرایند مکرر تقسیم‌بندی داده‌ها ممکن است مقدار بهینه واقعی به‌دست نیامده باشد. البته این مشکل اساساً شامل همهٔ روش‌های غیرخطی مثل شبکه‌های عصبی، الگوریتم ژنتیک، تحلیل ممیزی، و سایر روش‌های رده‌بندی نیز می‌شود. در مقابل، در رگرسیون لجستیک به‌دلیل اینکه متغیرهای مستقل از جنس پیوسته است، با اندکی تغییر، در جواب نهایی مقدار کمی تغییر حاصل می‌شود و می‌توان با استفاده از ترکیب این دو روش با یکدیگر سیستم رتبه‌بندی مناسب‌تری برای بازده سهام به‌دست آورد (ایمانی، ۱۳۹۰).

۶ روش ترکیبی کارت و رگرسیون لجستیک

در این مقاله از روش ترکیبی کارت و رگرسیون لجستیک استفاده خواهیم کرد. برای این منظور فرض کنید که n سهام برای دسته‌بندی در اختیار داریم. اگر فرض کنیم $X_{n'm}$ ماتریسی از m ویژگی شاخص برای هریک از n سهام باشد که هر ستون آن یکی از ویژگی‌های بارز در هر سهم، مثل نسبت ارزش دفتری به بازار (B/P) باشد که m_1 تا از آن‌ها ذاتاً متغیرهای پیوسته است و در طول زمان به‌صورت گسسته مشاهده شده‌اند و m_2 تا از آن‌ها به‌صورت ذاتی گسسته هستند (مثل نوع صنعت سهام). حال اگر فرض کنیم درخت تصمیم بهینه بر اساس داده‌های سهام متفاوت در نهایت منجر به K گره خواهد شد که هر گره n_k سهم وجود دارد که از این تعداد s_k تای آن سهام سودآور و مابقی $n_k - s_k$ سهم زیان‌ده است؛ بنابراین، اگر $P_k = \frac{s_k}{n_k}$ فراوانی نسبی سهام سودآور باشد. این نسبت در واقع توسط n_k سهم موجود در گرهٔ kام به‌دست آمده است و به‌طریقی، جمع‌بندی k سهم موجود در گره kام تلقی می‌شود. بنابراین می‌توان به کمک آن اگر سهم jام در گرهٔ kام قرار گرفته باشد، با استفاده از مدل‌بندی $p_{jk} = p_k + \varepsilon_{jk}, j = 1,\dots,n, k = 1,\dots,K$. احتمال دقیق‌تری را

[1] Zhu et al.

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

برای سودآور بودن سهم jام در گرهٔ kام به‌دست آورد. در این مدل e_{kj} ناشی از تفاوت احتمال سودآور بودن سهم kام را کنترل می‌کند.

در ادامه به‌منظور تکمیل مراحل الگوریتم کافی است احتمال‌های P_k را به‌گونه‌ای تغییر دهیم که بتوان بهترین مقدار P_{kj} را برای سهم j از آن به‌دست آورد. برای این منظور، مقادیرP_{jk} را با استفاده از مدل‌بندی رگرسیون لجستیک مدل‌بندی می‌کنیم. در این بخش از مدل‌بندی به‌دلیل ساختار رگرسیون لجستیک، تنها متغیرهای ذاتاً پیوسته قابل مدل‌سازی هستند. به‌عبارتی‌دیگر با استفاده از مدل:

$$\text{logit}(P_{jk}) = \text{logit}(P_k) + X_{j.}^{k} b \qquad (۱)$$

که در آن $X_{j.}^{k}$ سطر jام ماتریس اصلی داده‌های ذاتاً پیوسته است و بردار b به طول m_1 فی‌الواقع بردار ضرایب آن است، درصورتی‌که ضرایب آن معنی‌دار نباشد، بدین معنی است که همان احتمال‌های pk که از درخت تصمیم معمولی به‌دست می‌آید، توانایی کافی برای پیش‌بینی نسبت سهام سودآور در این مرحله از الگوریتم را داشته است و الگوریتم جدید کارایی لازم را ندارد (ژو و همکاران، ۲۰۱۲).

به‌طورکلی دو روش را برای پیاده‌سازی الگوریتم ترکیبی پیشنهاد می‌شود. در روش اول، در هر گره یک مدل رگرسیون لجستیک انجام می‌شود. یکی از معایب استفاده از این روش این است که رگرسیون لجستیک به‌صورت محلی و فقط برای داده‌های اختصاص‌یافته به آن گره باید انجام شود. لذا داشتن داده‌های بزرگ از ضرورت‌های انجام این روش است. در روش دوم در ابتدا و قبل از شروع گره اول، یک رگرسیون لجستیک بر اساس تمام داده‌ها انجام داده و درخت تصمیم را به اجرا می‌گذاریم. در این روش علی‌رغم اینکه نتایج گرهٔ اصلی تأثیر زیادی در کل نتایج خواهد داشت، می‌توان گفت از پایداری به‌نسبت بالاتری برخوردار است (همان منبع). در ادامهٔ این مقاله، از روش دوم به‌منظور استخراج نتایج استفاده شده است. یکی از مزیت‌های این روش این است که اگر اثر ویژگی خاصی در درخت تصمیم معمولی دیده نشده باشد، ممکن است در همین ابتدا اثر آن در رگرسیون لجستیک معنی‌دار باشد.

۷ پیاده‌سازی مدل و نتایج آن

در این بخش داده‌های مورداستفاده را با اختصار توضیح می‌دهیم و نتایج استفاده از مدل ترکیبی را بر روی داده‌ها را به‌دست می‌آوریم و در نهایت، با روش‌های دیگر مقایسه می‌کنیم.

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

کلیهٔ محاسبات و تحلیل‌های انجام‌شده در این مقاله با استفاده از نرم‌افزار R و بسته‌های جانبی پی. ال. ام.[1]، کورپلات[2] و آر-پارت، و ریم تری انجام شده است.

۷. ۱ داده‌ها، پیش‌پردازش، و انتخاب ویژگی‌ها

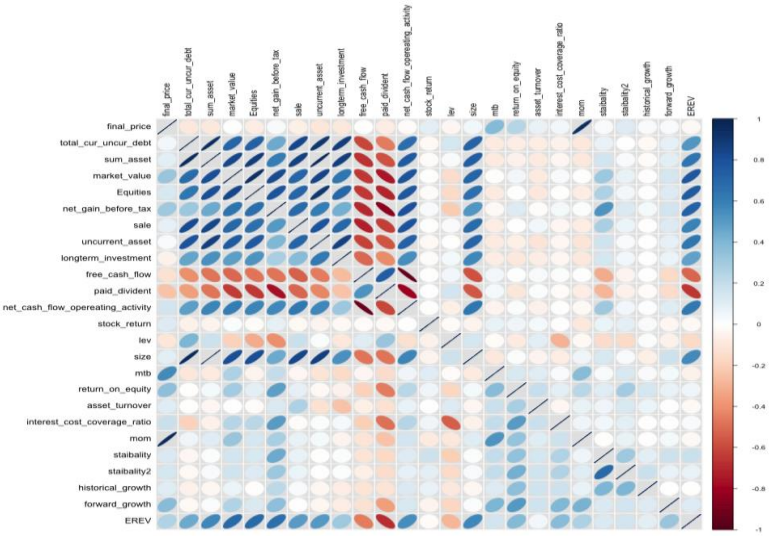
در اولین گام میزان بازده سه‌ماههٔ آتی را بر اساس آخرین قیمت در آخرین روز هر فصل موردمحاسبه قرار داده‌ایم. سپس یک متغیر جدید دوحالتی را بر اساس بازده سه‌ماههٔ آتی در هر بازهٔ زمانی به‌عنوان متغیر هدف در مدل در نظر گرفته شده است. مقادیر این متغیر جدید بدین صورت است که اگر بازده سهم بیشتر از میانهٔ بقیهٔ سهام در همان مقطع زمانی باشد، مقدار ۱ و در غیر این صورت، مقدار صفر را خواهد گرفت. این متغیر در واقع به‌عنوان متغیر وابسته در مدل‌سازی الگوریتم طبقه‌بندی پیشنهادی ما در نظر گرفته می‌شود. به عبارت دقیق‌تر، به کمک این متغیر می‌توانیم سهامی را که درصد سودآوری آن بیشتر از متوسط سود سهام در بازار متعادل است، در طول مدت سه ماه تعیین کنیم. مزیت استفاده از این متغیر دوحالتی در بازهٔ زمانی سه‌ماهه نسبت‌به حالت پیوسته که به‌صورت روزانه به‌دنبال پیش‌بینی خواهیم بود، این است که شرکت‌هایی را انتخاب خواهیم کرد که به‌طور متوسط در مدت‌زمان بیشتر سودآور هستند و در واقع، به‌دنبال سهامی نیستیم که در مقطع خاصی از زمان فقط سودآور هستند.

به‌منظور انتخاب ویژگی مناسب از بین ویژگی‌های موجود، بر اساس نظریه‌ها و مقالات انجام‌شده و همچنین فرضیهٔ استقلال متغیرهای پیش‌بینی‌کننده در مدل، مجموعه‌ای از ویژگی‌های ترکیبی ساخته شده است. برای بررسی این موضوع یکی از اصلی‌ترین فرضیاتی که در مدل‌های رگرسیونی وجود دارد نبود وابستگی متغیرهای مستقل به یکدیگر است. ازآنجاکه ضریب همبستگی پیرسون در داده‌های سری زمانی کارایی لازم را ندارند، در شکل ۱ میزان ضریب همبستگی مقطعی زمانی را برای هریک از متغیرهای مربوط به شرکت‌ها را استخراج کرده‌ایم. مقادیر همبستگی‌های نزدیک به ۱ و ۱- به‌صورت تقریباً خطی با شیب مثبت و منفی و مقادیر نزدیک به صفر با شکل‌های بیضی‌گون گزارش شده است. هرچه میزان همبستگی به صفر نزدیک شود، رنگ موجود در نمودار نیز کم‌رنگ‌تر شده است. همچنین با استفاده از روش دسته‌بندی k- میانگین برای ماتریس همبستگی تشکیل‌شده و مرتب‌سازی

[1] Plm
[2] Corrplot

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

ماتریس نهایی، متغیرهایی که بیشترین همبستگی را با یکدیگر دارند، در کنار یکدیگر قرار داده‌ایم تا بتوان متغیرهای همبسته را از متغیرهای ناهمبسته به‌راحتی تشخیص داد.



شکل ۱. میزان ضریب همبستگی مقطعی (زمانی) فصلی ویژگی‌های ۱۰۰ شرکت در بازهٔ زمانی سال‌های ۱۳۹۰ تا ۱۳۹۷

منبع: محسبات محقق

همان‌طور که در شکل ۱ مشخص است، تعداد زیادی از متغیرها رابطهٔ خطی زیادی با یکدیگر دارند که نشان از انتخاب ویژگی‌هایی با اطلاعات مشابه می‌دهد. همان‌طور که می‌دانیم در مدل‌های پارامتری اصولاً بایستی متغیرهای پیشگو از یکدیگر مستقل باشند. اما به‌دلیل غیرپارامتری بودن، مدل غیرخطی در نظرگرفته شده است. لذا در این مقاله، در عمل نگرانی از این بابت وجود ندارد. بااین‌وجود به‌منظور شناسایی مطابقت مدل پیشنهادی با متغیرهای ترکیبی پیشنهادی در مقالات دیگر و به‌دلیل پیش‌پردازش داده‌ها که مشابه آن در تمام روش‌های مبتنی بر هوش مصنوعی انجام می‌شود، نیاز است ویژگی‌های ترکیبی از متغیرهای همبسته ساخته شود. در واقع، ویژگی‌های جدید میانگین وزنی از متغیرهای اصلی است که به کمک بردار امتیازات در اولین مؤلفه در تحلیل مؤلفه‌های اصلی ساخته شده‌اند. همان‌طور که می‌دانیم، مؤلفه‌های اصلی در واقع بهترین ترکیب خطی است که می‌تواند بیشترین تغییرات (واریانس) چند ویژگی خاص را به کمک آن نشان داد.

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

اولین ویژگی میانگین وزنی از ارزش‌های مالی هر سهم شامل ۱ـ سود سهام به قیمت، گردش وجوه نقدی (مجموع وجوه نقدی آزاد و ناشی از فعالیت‌های عملیاتی) به قیمت، میزان فروش به قیمت، و ارزش دفتری به قیمت (VAL)، ۲ـ میانگین وزنی از سودهای شرکت شامل بازده حقوق صاحبان سهام، بازده نقدی صاحبان سهام، و گردش دارایی‌ها (PROFIT)، ۳ـ متغیر میانگین وزنی از بدهی‌های جاری و غیرجاری به حقوق صاحبان سهام و بدهی جاری و غیرجاری به ارزش بازار (LEV)، ۴ـ میانگین وزنی نسبت پوشش هزینهٔ بهره و نسبت جریان نقد آزاد به بدهی (DEBT)، و ۵ـ میانگین مومنتوم‌های[1] شش‌ماهه و سالیانه است (MOM).

همچنین ویژگی‌های خاص دیگری که در مقالهٔ ژو و همکاران (۲۰۱۲) به‌منظور پیش‌بینی بازده آتی وارد مدل‌سازی آماری شده است. این ویژگی‌ها عبارت‌اند از ۱ـ میزان پایداری[2] سهم که شاخصی ترکیبی از میانگین وزنی یکسان نوسانات درآمد و فروش و جریان پول نقد در پنج سال گذشته است، ۲ـ ویژگی رشد تاریخی[3] که عبارت است از میانگین وزنی رشد سود، فروش، و جریان وجوه نقد در سه سال گذشته است، ۳ـ ویژگی رشد آینده[4] نیز که عبارت است از میانگین وزنی انتظارات از رشد سود پیش‌بینی‌شده برای سال مالی ۱ و ۲، و ۴ـ سود تجدیدنظر روی سود هر سهم[5] که میانگین وزنی تغییرات سه‌ماهه در پیش‌بینی انتظارات سود برای سال مالی ۱ و ۲ که هر دوی آن‌ها از پایگاه داده‌ای I/B/E/S به‌دست آمده، نیز در مدل‌سازی رگرسیونی پانلی درخت تصمیم وارد شده است.

پس از استخراج متغیرهای جدید از داده‌های خام، به‌جای استفاده از مقادیر اصلی، ویژگی‌ها از رتبهٔ آن‌ها در هر مقطع زمانی در بین n سهام که بر n تقسیم شده است، استفاده می‌کنیم؛ بنابراین، تمام ویژگی‌های اعداد به عددی در بازهٔ صفر و ۱ تبدیل خواهد شد. دلیل اصلی این موضوع حساسیت بالای درخت تصمیم به مقادیر دور افتاده است که استفاده از داده‌های رتبه‌بندی‌شده از حساسیت نتایج نسبت‌به داده‌های دورافتاده کم می‌کند. زیرا در آن صورت مقادیر داده‌ها مهم نیست و رتبهٔ آن‌ها در مقایسه با دیگر سهام ملاک قرار داده می‌شود. در نهایت ماتریس همبستگی مقطعی زمانی ویژگی‌های نهایی در جدول ۱ گزارش

[1] Momentum
[2] Stable amount (STAB)
[3] Historical growth (HISTGR)
[4] Future growth (FORWGR)
[5] Profit revision on earnings per stock (EREV)

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

شده است. همان‌طور که انتظار می‌رفت میزان همبستگی در اکثر ویژگی‌ها در بازهٔ ۰/۲ تا ۰/۲ــ شده است.

جدول ۱ضریب همبستگی مقطعی زمانی ویژگی‌های نهایی

	PROFIT	LEV	DEBT	STAB	HISTGR	FORWGR
PROFIT	۱	ــ۰/۰۳۱	۰/۱۹۷	۰/۲۶۲	۰/۲۵۹	۰/۱۸
LEV	۰/۰۳۰۱	۱	ــ۰/۰۵	ــ۰/۲۱۹	ــ۰/۱۴۳	ــ۰/۲۳n
DEBT	۰/۲۰۳	۰/۹۵	۱	۰/۲۳۲	۰/۱۵۶	۰/۴۲۱
STAB	۰/۲۶۲	ــ۰/۲۱۹	۰/۲۳۲	۱	۰/۱۶۳	۰/۲۲۴
HISTGR	۰/۲۵۹	۱۴۶۰	۰/۱۵۶	۰/۱۶۳	۱	۰/۱۵۷
FORWGR	۰/۱۸	ــ۰/۲۳	۰/۴۲۱	۰/۲۲۴	۰/۱۵۷	۱

منبع: محاسبات محقق

جدول ۲

مقدار آمارهٔ t برای شیب خط رگرسیونی هریک از متغیرهای ترکیبی به‌صورت جداگانه و میزان بازدهی سالیانه بر اساس آن‌ها

بازده سالانه (٪)				
فاکتور مرکب	**آمارهٔ t**	**زیاد**	**متوسط**	**کم**
VAL	۲/۷۸	۱۰/۱۱	۹/۸	۸/۵
PROFIT	۲/۱۳	۱۱/۲۵	۱۰/۲۷	۹/۱۵
LEV	۱/۷۸	۹/۸۲	۸/۹۲	۹/۱۴
STAB	۰/۸۹	۵/۵۲	۴/۳۸	۶/۷۵
FORWGR	ــ۰/۰۲	۶/۴۱	۵/۱۷	۸/۹۸
HISTGR	۰/۱۹	۷/۶۵	۸/۲۳	۵/۱۴
DEBT	۱/۰۳	۱۰/۰۴	۹/۱۱	۸/۱۲

منبع: محاسبات محقق

۲.۷ مدل‌سازی بازده سهام

۱.۲.۷ پیش‌بینی بازده و انتخاب پرتفوی

اصولاً در مدل‌سازی ضروری است که به‌صورت جداگانه به کمک تجزیه‌وتحلیل هریک از متغیرهای ترکیبی به بینش مختصری برای هریک از آن‌ها دست یابیم. برای این منظور، ما

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

بازده‌های مازاد آتی یک‌ماهه را بر اساس هریک از فاکتورهای ترکیبی، رگرسیون گرفته و آمارهٔ t ضریب زاویهٔ شیب خط را گزارش کردیم. به علاوه، فاکتورهای مرکب را در سه گروه پرتفوی سهام با اندازهٔ برابر افراز کردیم. لذا هر سه ماه، یک پرتفوی «زیاد»، «متوسط»، و «کم» تشکیل می‌شود و بر اساس آن، میزان بازده سالیانهٔ هر سبد را استخراج کرده‌ایم. لازم به ذکر است، هزینه‌های کارمزد خریدوفروش در این سبدها صفر در نظرگرفته شده است. جدول ۲ آمارهٔ t یک‌متغیره، بازده‌های سالانهٔ این سه پرتفوی طی کل مدت نمونه، و بازده پرتفوی زیاد_کم را نشان می‌دهد. تجزیه‌وتحلیل یک‌متغیره داده‌ها پیشنهاد می‌کند که ارزش، سودآوری، و بازنگری‌های عایدات مهم‌ترین عوامل تعیین‌کنندهٔ بازده‌های آتی است که اهرم مالی و مومنتم تاریخی پس از این عوامل قرار می‌گیرد.

به‌منظور بررسی دقیق‌تر روش پیشنهادی ما برای پیش‌بینی بازده سهام، داده‌های فصلی از سال ۹۰ تا ۹۶ در واقع، پیش‌بینی یک فصل جلوتر را انجام می‌دهد و سپس می‌توانیم بر اساس آن، سهام شرکت‌های مناسب را در ابتدای هر فصل تشخیص دهیم و سبد سهامی را پیشنهاد دهیم که به‌طور متوسط سودآورتر از سهام دیگر بازار است. بر اساس این پیش‌بینی، پرتفوی‌هایی را پیشنهاد داده‌ایم و کارایی آن را با توجه به داده‌های خارج از مدل یادگیری (سال ۹۷) می‌سنجیم. به‌طور خاص، سه پرتفوی بدین شکل تشکیل می‌شود: پرتفوی اول (p_1) که محتوی یک‌سوم از سهام است؛ سهامی که دارای کمترین احتمال عملکرد بهتر است (یعنی پرتفوی عملکرد بدتر)؛ پرتفوی دوم (p_2) که محتوی یک‌سوم بعدی است (یعنی پرتفوی بازار)، و پرتفوی سوم (p_3) که محتوی یک‌سوم باقیمانده از سهام است؛ سهامی که دارای بیشترین احتمال عملکرد بهتر است. ما برای ساخت پرتفوی از طرح وزن‌دهی یکسان استفاده کردیم. پرتفوی‌ها را هر سه ماه یک‌بار مجدداً موازنه و سودها را مجدداً محاسبه می‌کنیم. برای سهولت، هزینه‌های تراکنش و کارمزد خریدوفروش در ارزیابی عملکرد پرتفوی‌ها صفر در نظرگرفته شده است.

به‌منظور ارزیابی هریک از پرتفوی‌ها، چندین عامل را برای هر پرتفوی موردبررسی قرار می‌دهیم. یکی از این روش‌ها محاسبهٔ ریسک تعدیل‌شده بازده اضافی ناشی از سهام موجود در پرتفوی است. برای این منظور، چند مجموعه فاکتور را در نظر می‌گیریم. اولین عامل مدل سه عاملی فاما و فرنچ (۱۹۹۳) است که در محاسبهٔ آن علاوه بر فاکتور مدل قیمت‌گذاری

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

دارایی‌های سرمایه‌ای[1] دو عامل ریسک سرمایه شرکت[2] و نسبت ارزش دفتری به ارزش بازار[3] نیز در نظر گرفته می‌شود. در این روش بر اساس رابطه:

$$r = R_f + b(K_m - R_f) + b_s \times SMB + b_v \times HML + a \qquad (۲)$$

که در آن r: نرخ بازده موردانتظار پرتفوی، R_f : نرخ بازده بدون ریسک، K_m : بازده پرتفوی بازار، و β سه‌عاملی، مشابه همان β کلاسیک است اما با آن برابر نیست. زیرا الآن دو عامل اضافهٔ دیگر به اسم SMB و HML نیز به مدل اضافه شده است که به‌ترتیب، مشخص‌کنندهٔ بزرگی شرکت، و بازده مازاد سهام ارزشی نسبت‌به سهام رشدی را اندازه‌گیری می‌کند. این عوامل، با ترکیب پرتفوی‌هایی متشکل از سهام رتبه‌بندی شده (رتبه‌بندی B/M و رتبه‌بندی سرمایه) و داده‌های تاریخی بازار در دسترس، محاسبه می‌شود. سپس با رگرسیون خطی بر روی فاکتورهای ریسک به‌صورت

$$r_B - r_B = a + a_{m-1}^{L} J_m b_m \qquad (۳)$$

که در آن r_p بازده پرتفوی و r_B بازده کل سهام موجود در نمونه است و b_m ضریب پرتفوی p بر روی فاکتور ریسک m بوده است.

هدف اصلی مطالعه بررسی روش ترکیبی معرفی‌شده در مقایسه با روش کارت و رگرسیون لجستیک به‌تنهایی است. برای این منظور، میزان بازدهی مربوط به نمونه‌های سال ۱۳۹۷ استخراج و توسط هریک از روش‌ها به‌دست آورده می‌شود. همچنین به‌منظور افزایش کیفیت نتایج، از تکنیک‌های باز نمونه‌گیری به کمک الگوریتم درخت‌های جنگل تصادفی (بریمن، ۲۰۰۱) نیز استفاده کردیم که بتوانیم در نهایت خروجی مطلوب‌تری را ارائه دهیم. در این روش به‌جای به‌دست آوردن یک درخت تصمیم، به تعداد زیادی این عمل را تکرار کرده و در نهایت خروجی مدل را از میانگین نتایج کلیهٔ درخت‌های نمونه به‌دست می‌آوریم. اگرچه روش جنگل تصادفی را می‌توان به‌راحتی در همهٔ انواع داده‌ها استفاده کرد و پایداری مناسبی به داده‌ها می‌دهد، عدم تفسیرپذیری برخی نتایج را می‌توان از معایب آن دانست. درخت جنگل تصادفی در مواردی مانند داده‌های با خطای زیاد، بیش‌برآوردی پارامترهای مدل را نتیجه می‌دهد.

---

[1] Capital Asset Pricing Model (CAPM)

[2] High Minus Low, HML

[3] Small Minus Big, SMB

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

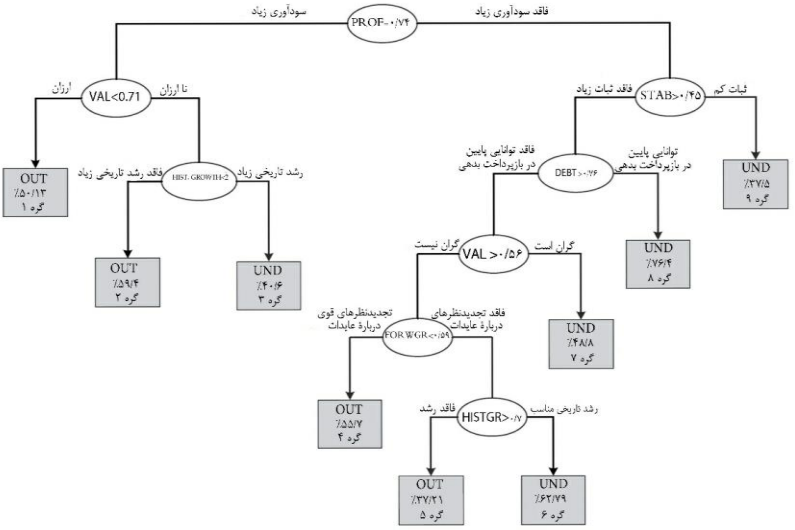
۲.۲.۷ نتایج الگوریتم ترکیبی

در پایان هر سه ماه، شش فاکتور مرکب را در سه بلوک با وزن یکسان تقسیم‌بندی کردیم. با استفاده از داده‌های آموزشی[1] یک مدل کارت ساختیم و با پیش‌بینی داده‌ها، مدل‌های بعدی را در پایان هر سه ماه تا سال ۱۳۹۷ به‌دست آوردیم. سپس از یک رگرسیون لجستیک به‌منظور تنظیم احتمالات مبتنی بر مدل کارت استفاده کردیم. برخلافِ روش کارت، رگرسیون لجستیک نمی‌تواند فاکتورهای بی‌معنی را حذف کند و لذا از روشِ گزینش متغیرـ AIC (معیار آکائیک) به‌منظور انتخاب متغیرهای مهم استفاده کردیم.

مثالی از درخت‌هایی که در مدل داده‌های آموزشی تشکیل دادیم، در شکل ۲ رسم شده است. همان‌گونه که نشان داده شده است، شانس یک سهام که بازدهی بالاتر از بازده مرجع تولید کند، یعنی بازده‌های با وزن یکسان سهام تدافعی، اصولاً توسط آثار توأم سودآوری، ارزش، و ثبات تعیین می‌شود. اگرچه سایر فاکتورها در سطوح نسبتاً پایین‌تر سلسله‌مراتب تصمیم است، در واقع، اولین مشاهده‌ای که باید به آن توجه کرد این است که اولین تقسیم بر اساس سودآوری صورت می‌گیرد و به‌طور خاص‌تر یعنی تمایز میان سهامی که سودآور است (شاخهٔ سمت چپ) و سهامی که خیلی سودآور نیست. به‌طورکلی، مدل درخت سهام ارزان‌قیمت و سودآور (گرهٔ ۱) را ترجیح می‌دهد و احتمال ۵۰/۱۳ را برای بهتر عمل کردن نسبت به کل جامعهٔ آماری را به این سهام نسبت می‌دهد. علاوه بر معیار سودآوری کم، سهامی که بدتر از سهام مرجع عمل می‌کند دارای یک‌سری ویژگی‌ها نظیر ثبات کم (گرهٔ ۹) و توانایی ضعیف برای بازپرداخت بدهی‌هاست (گرهٔ ۸).

[1] Educational data

[DOR: 20.1001.1.26453355.1399.13.45.2.5]



شکل ۲. درخت تصمیم در اسفندماه ۱۳۹۵.
منبع: محاسبات محقق

احتمال‌های عملکرد بهتر در گره‌های پایانی گزارش شده است. OUT دلالت دارد بر عملکرد بهتر و UND دلالت دارد بر عملکرد بدتر. در هر دوره، با استفاده از سه قالب، فاکتورهای گروه را با وزن یکسان تقسیم کردیم.

جدول ۳ ضریب‌های برآوردشده از رگرسیون لجستیک در اسفندماه ۱۳۹۷ را نشان می‌دهد. ما از روش گام‌به‌گام گزینش متغیر_AIC استفاده کردیم تا متغیرهای مهم را گزینش کنیم. مدل نهایی دارای سه فاکتور مرکب بود.

جدول ۳. ضریب‌های برآورد شده از رگرسیون لوجستیک در اسفندماه ۱۳۹۷

متغیر	ضریب	P-Value
DEBT	۰/۰۵۴۰	۰/۰۱۱**
VAL	۰/۰۳۱	۰/۰۰۵**
PROF	۰/۰۷۸	۰/۰۷۳ *

*نشان‌دهندهٔ معناداری به میزان ۱۰٪ است.
**نشان‌دهندهٔ معناداری به میزان ۵٪ است. (منبع: محاسبات محقق)

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

البته استثنائاتی در قاعدهٔ کلی و عمومی وجود دارد که یکی از فواید و مزیت‌های روش‌های مبتنی بر درخت است. برای مثال، سهامی که به‌صورت جذاب قیمت‌گذاری شده است، ثبات بالایی دارد، و توانایی کافی برای بازپرداخت بدهی‌هایشان را دارد، بهتر از مرجع عمل می‌کند. به‌ویژه اگر دارای تجدیدنظرهای قوی در زمینهٔ عایدات باشد (گرهٔ ۴) یا رشد تاریخی بالایی داشته باشد (گرهٔ پنج) و حتی اگر سودآوری آن‌ها ضعیف باشد. به علاوه، فاکتور مرکب ارزش در دو گرهٔ تقسیم‌کنندهٔ مختلف با مقادیر تقسیم‌کنندگی متفاوت اعمال شده است. تعامل میان ارزش و سایر فاکتورهای مرکب، به‌سادگی از طریق ساختار سلسله‌مراتبی درخت به‌دست می‌آید.

مرحلهٔ دوم در روش ترکیبی این است که از رگرسیون لجستیک استفاده و احتمال‌های تولیدشده توسط مدل کارت را تنظیم کنیم. این مرحله این توانایی را دارد که تأثیرات خطی فاکتور و هر نوع تأثیر فراموضعی را که مدل درخت ممکن است از قلم انداخته باشد، مشارکت دهد و همچنین رویهٔ پاسخ را که تا حدی هموارتر است، تولید کند. نتایج تجزیه‌وتحلیل ریسک برای پرتفوی‌های P_1 و P_3 در جدول ۴ به‌طور خلاصه شرح داده شده است. در سطح معناداری ۵٪ ، بازده‌های تعدیل‌شده از لحاظ ریسک پرتفوی‌های بلند و کوتاه برای هر دو مجموعه از فاکتورها، از لحاظ آماری معنادار است.

جدول ۴. نتایج تجزیه و تحلیل ریسک پرتفوی‌های P_1 و P_3

	P_1	P_3
a	-۰/۰۲۲	۰/۰۲۴
	(۰/۰۱۴)**	(۰/۰۰۷)***
MKT	۰/۲۱۴	-۰/۱۷۳
	(۰/۰۰۷)***	(۰/۰۴)**
SMB	۰/۰۸۷	-۰/۰۹۰
	(۰/۰۵۵)	(۰/۰۱۲)**
HML	-۰/۳۸۱	۰/۲۰۵
	(۰/۰۲۱)***	(۰/۰۰۸)**

** نشان‌دهندهٔ معناداری به میزان ۵٪ است.

*** نشان‌دهندهٔ معناداری به میزان ۱٪ است. (منبع: محاسبات محقق)

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

جدول ۴ ضرایب پیشگویی‌شده توسط رگرسیون لجستیک در تاریخ اسفند ۱۳۹۷ را نشان می‌دهد. مقدم بر احتمال‌های پیشگویی‌شده توسط درخت، رگرسیون لجستیک، وزن‌های مربوط به سودآوری را روبه‌پایین تنظیم می‌کند (یعنی دارای ضریب منفی است)، درحالی‌که تأکید روی تجدیدنظرها دربارهٔ عایدات و ارزش را افزایش می‌دهد. این مسئله عنوان می‌کند که مدل کارت تاحدزیادی تحت تأثیر سودآوری قرار می‌گیرد که اولین قاعدهٔ تقسیم است؛ اما به‌اندازهٔ کافی تأثیرات خطی را که به‌ویژه از نظریات تحلیلگران حاصل می‌شود، دربرنمی‌گیرد.

۳. ۲. ۷ مقایسه مدل‌ها

اگرچه اجرای روش دو ترکیبی روی سهام تدافعی نتیجه‌بخش بود، به‌منظور مقایسهٔ کیفی آن با برخی از روش‌های رقیب آن را مقایسه می‌کنیم. این روش‌ها عبارت‌اند از مدل کارت، مدل لجستیک (که هریک به‌طور جداگانه برآورد شده است)، و روش جنگل تصادفی که یک‌بار دیگر از فرایند گزینش متغیر_AIC استفاده می‌کنیم تا متغیرهای مهم در تجزیه‌وتحلیل رگرسیون لجستیک را انتخاب کنیم. پس از انجام دادن مراحل فوق، کارایی سه مدل را برحسب توانایی‌شان در گزینش سهام، به‌صورت خارج از نمونه موردارزیابی قرار می‌دهیم. به‌طور خاص، هر ماه تمام سهام را بر مبنای مقادیر برازش‌شده را که توسط سه مدل فوق از تاریخ فروردین ۱۳۹۰ تا زمستان ۱۳۹۷ ارائه می‌شد، رتبه‌بندی کردیم. بعدازآن، از استراتژی پرتفوی استفاده کردیم تا مدل‌ها را خارج از نمونه مقایسه کنیم. برای این کار، هرماه سهام را بر اساس احتمال‌های پیش‌گویی‌شده برای عملکرد بهتر مانند قبل به سه گروه مساوی تقسیم کردیم و بدین‌ترتیب، سه پرتفوی را تشکیل دادیم و آماره‌های زیر را برای آن‌ها استخراج کردیم. نرخ اصابت سری‌های زمانی: این نرخ به‌صورت $\frac{T^+}{T}$ محاسبه شد که T^+ تعداد ماه‌ها با بازده‌های مرجع در مدت خارج از نمونه است و T تعداد ماه‌ها در مدت خارج از نمونه است. نرخ اصابت سری‌های زمانی عبارت است از نسبت ماه‌هایی که پرتفوی بهتر از مرجع عمل کرده است.

۱. بازده مازاد: بازده مازاد عبارت است از بازده مازاد پرتفوی که به‌صورت سالانه محاسبه شده است بر روی بازده مرجع.

$$r_e = [(\prod_{t=1}^{T}(1+r_{pt}) - \prod_{t=1}^{T}(1+r_{bt}))^{1/T}]^{12freq} \qquad (۴)$$

که r_{pt} و r_{bt} به‌ترتیب بازده‌های پرتفوی و بازده مرجع در زمان t هستند. $\prod_{t=1}^{T}(1+r_{pt})$ و $\prod_{t=1}^{T}(1+r_{bt})$ بازده‌های تجمعی در طول دوره‌های T برای پرتفوی و

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

بازده‌های تجمعی در طول دوره‌های T برای مرجع است. توان $12/freq$ برای سالانه کردن بازده است که $freq$ عبارت است از فراوانی در ماهی که بازده‌ها ثبت می‌شود. در این تحقیق، بازده در فراوانی سه‌ماهه است؛ بنابراین، $freq = 3$ است.

۲. خطای ردیابی: خطای ردیابی به‌صورت زیر سالانه شده و محاسبه می‌شود:

$$s = \sqrt{\frac{۱۲}{freq}}\, sd\left(r_{pt} - r_{bt}\right) \qquad (۵)$$

که sd انحراف استاندارد است و همانند قبل $freq$ عبارت است از فراوانی در ماهی که بازده‌ها ثبت می‌شود.

۳. نسبت اطلاعات: نسبت اطلاعات عبارت است از میانگین سالانه‌شدهٔ تفاضل میان بازده پرتفوی و بازده مرجع که بر خطای ردیابی سالانه‌شده تقسیم می‌شود (محاسبهٔ خطای ردیابی در بالا شرح داده شد).

$$IR = \frac{12/freq \sum_{t=1}^{T}(r_{pt}-r_{bt})}{\sqrt{\frac{۱۲}{freq}} sd(r_{pt}-r_{bt})} = \sqrt{\frac{12}{freq}} \frac{\sum_{t=1}^{T}(r_{pt}-r_{bt})}{sd(r_{pt}-r_{bt})} \qquad (۶)$$

۴. معیار شارپ: عبارت است از میانگین سالانه‌شدهٔ تفاضل میان بازده پرتفوی و بازده بدون ریسک (بازده مرجع) است که بر خطای استاندارد بازده‌های سهام موجود در هر پرتفوی تقسیم می‌شود:

$$SI = \frac{r_{pt}-r_{bt}}{Sd} \qquad (۷)$$

۵. دورهٔ نگهداری سهام: معیاری برای گردش پرتفوی است. هرچه مدت نگه‌داشتن سهام بیشتر باشد، گردش پرتفوی و هزینه‌های تراکنش کمتر خواهد بود. دورهٔ نگهداری یک سهام در یک پرتفوی عبارت است از میانگین دوره‌های نگهداری در نمونه خارج از آزمون. برای مثال، سهام A برای سه دورهٔ متوالی در یک پرتفوی می‌ماند و سپس از آن پرتفوی خارج می‌شود. سپس دوباره به آن پرتفوی برمی‌گردد و برای دو دورهٔ متوالی در آن پرتفوی می‌ماند. دورهٔ نگهداری سهام A در این پرتفوی مساوی است با $n = ۲.۵$ یعنی میانگین دوره‌های نگهداری سهام A در این پرتفوی. برای هر پرتفوی، اجازه دهید n تعداد سهامی باشد که در پس‌آزمایی همواره وارد پرتفوی می‌شود و $h_1, h_2, \ldots, h_n$ سری‌های دورهٔ نگهداری متناظر با این سهام باشد. ما میانگین و میانهٔ دورهٔ نگهداری را بر اساس سری‌های دوره نگهداری پرتفوی، گزارش کرده‌ایم.

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

جدول ۵ نتایج نرخ اصابت پانلی خارج از نمونه، نرخ اصابت سری‌های زمانی، نسبت اطلاعات، بازده‌های مازاد، معیار شارپ، و اطلاعات مربوط به دورهٔ نگهداری سهام را به‌صورت خلاصه نشان می‌دهد؛ برای پرتفوی‌هایی که با استفاده از چهار روش یادشده ساخته شده‌اند. می‌توانیم مشاهده کنیم که اگرچه روش کارت نسبت‌به مدل رگرسیون لجستیک و مدل جنگل تصادفی بهتر عمل می‌کند، عملکرد پرتفوی‌های کوتاه و بلندی که توسط روش هیبریدی ساخته شده‌اند، در این دورهٔ زمانی که به‌عنوان نمونه در نظر گرفته‌ایم، نسبت‌به بقیه عملکردها برتر و بالاتر است. صرف‌نظر از روش‌های تشخیص عملکرد، این موضوع صحت دارد که عملکرد پرتفوی‌هایی که از روش دوهیبریدی به‌دست می‌آیند، نسبت‌به عملکرد پرتفوی‌هایی که از سایر روش‌ها به‌دست می‌آیند بهتر است. به علاوه همان‌گونه که از مقادیر نرخ اصابت سری‌های زمانی پیداست، بازده‌های بهبودیافته‌ای که با استفاده از روش دورگه به‌دست می‌آید، از سازگاری بالایی برخوردار است.

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

جدول ۵. نتایج پس‌آزمایی چهار مدل معرفی‌شده

مدل	پرتفوی	نرخ اصابت پانلی	نرخ اصابت TS	بازده مازاد (٪)	معیار شارپ SI	دورهٔ نگهداری	
						میانگین	میانه
جنگل تصادفی	P_1	۰۰/۴۲۴	۰۰/۴۷۸	۰۰/۷۸	۰/۶۱	۱/۵۱	۱/۵۵
	P_2	۰۰/۴۵۷	۰۰/۴۱۲	-۰۰/۱۵	۰/۶۴	۱/۳۶	۱/۲۸
	P_3	۰۰/۵۱۹	۰۰/۶۷۱	۱۰/۱۲	۰/۷۱	۱/۸۷	۱/۷۹
رگرسیون لجستیک	P_1	۰۰/۳۴۵	۰۰/۳۷۸	-۵/۱۴	۰/۵۶	۲/۱۴	۲/۱۹
	P_2	۰۰/۴۹۱	۰۰/۵۹۶	۱۰/۹۲	۰/۵۱	۲/۹۳	۲/۶۹
	P_3	۰۰/۶۴۳	۰/۸۹۱	۲/۵۶	۰/۳۸	۳/۱۶	۲/۸۹
کارت	P_1	۰۰/۳۳۲	۰۰/۳۶۰	-۴۳۲۲	۰/۷۲	۱/۱۸	۱/۱۱
	P_2	۰/۵۱۰	۰/۰۵۶	۱/۳۷	۰/۶۲	۲/۰۷	۱/۹۸
	P_3	۰/۵۷۲	۰۰/۷۰۲	۲/۱۹	۰/۷۳	۳/۸۷	۳/۰۸
روش هیبریدی	P_1	۰/۵۵۱	۰۰/۳۷۲	-۵/۰۴	۰/۶۳	۲/۲۶	۲/۲۵
	P_2	۰/۵۲۴	۰/۰۵۴۹	۱/۱۳	۰/۶۶۱	۲/۸۸	۲/۶۰
	P_3	۰/۶۰۳	۰۰/۷۱۲	۲/۰۸	۰/۸۳۲	۴/۰۱	۳/۹۸

منبع: محاسبات محقق

جدول ۵ نتایج پس آزمایی را برای چهار مدل معرفی‌شده گزارش می‌نماید. معیارهای سنجش عملکرد که در این پس‌آزمون مورداستفاده قرار گرفته‌اند، عبارت‌اند از؛ نرخ اصابت پانلی، نرخ‌های اصابت سری‌های زمانی (نرخ اصابت TS)، بازده‌های مازاد سالانه شده، معیار شارپ، و میانه و میانگین دورهٔ نگهداری سهام.

در رابطه با دورهٔ نگهداری سهام باید به این نکته توجه کرد که سهامی که در پرتفوی‌های P_1 و P_3 قرار دارد نسبت‌به سهامی که در پرتفوی P_2 قرار دارد، تمایل کمتری برای خریدوفروش دارد. به‌طورکلی، پرتفوی‌های مبتنی بر روش جنگل تصادفی، دست‌به‌دست شدن زیادی از خود نشان می‌دهند (یعنی آن‌ها دارای کمترین دورهٔ نگهداری سهام هستند).

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

همان‌گونه که توسط فاکتور میانهٔ دورهٔ نگهداری سهام نشان داده شده است، روش ترکیبی و رگرسیون لجستیک نسبت‌به دو روش دیگر دارای کمترین میزان دست‌به‌دست شدن سهام است. دلیل این ویژگی روش رگرسیون لجستیک این است که رویهٔ احتمالی که این روش تولید می‌کند هموار است و تغییرات در آن تدریجی‌تر.

طول دورهٔ نگهداری سهام در روش هیبریدی مساوی با طول دورهٔ نگهداری سهام در رگرسیون لجستیک است؛ بنابراین، عملکرد ارتقایافتهٔ پرتفوی‌هایی که توسط روش هیبریدی ساخته شده‌اند به نسبت عملکرد پرتفوی‌هایی که توسط روش‌های دیگر ساخته شده‌اند، براثرِ هزینه‌های تراکنش از بین نمی‌رود.

۸ جمع‌بندی و نتیجه‌گیری

در این پژوهش، با تمرکز بر مسئلهٔ پیش‌بینی‌پذیری سهام و مفاهیم مربوط به آن به بررسی روش پیشنهادی جدید برای پیش‌بینی بازده سهام و انتخاب پرتفوی مناسب بر اساس آن پرداخته شد. با توجه به نتایج حاصل از تحقیق، روش کارت نسبت‌به مدل رگرسیون لجستیک و مدل جنگل تصادفی بهتر عمل می‌کند، اما عملکرد پرتفوی‌های کوتاه و بلندی که توسط روش هیبریدی ساخته شده است، در دورهٔ زمانی تحقیق حاضر، نسبت‌به بقیهٔ عملکردها، برتر و بالاتر است. لذا عملکرد پرتفوی‌هایی که از روش ترکیبی به‌دست می‌آید، نسبت‌به عملکرد پرتفوی‌هایی که از سایر روش‌ها به‌دست می‌آید، بهتر و کاراتر است. لذا همان‌گونه که از مقادیر نرخ اصابت سری‌های زمانی پیداست، بازده‌های بهبودیافته‌ای که با استفاده از روش دورگه به‌دست می‌آید، از سازگاری بالایی برخوردار است. بنابراین، ازآنجاکه در روش‌های قبلی عموماً انتخاب ویژگی‌ها به‌صورت مستقل از مدل انجام می‌شود، یکی از مزیت‌های روش پیشنهادی ما این بود که به‌خوبی توانستیم برخی از ویژگی‌های رگرسیون لجستیک را هم‌زمان با ویژگی‌های خوب سادگی و قابل‌فهم بودن درخت تصمیم در اختیار داشته باشیم.

اجرای این روش سبب شناسایی مهم‌ترین متغیرهای ترکیبی VAL، DEBT، و PROF شد که می‌توانند نقش اساسی در انتخاب پرتفوی آتی داشته باشند. این متغیرهای ترکیبی می‌توانند در مدل‌هایی که وظیفهٔ انتخاب متغیرها را بر عهده ندارند، به‌خوبی جایگزین متغیرهای ورودی شوند. به علاوه سبدهای سهامی که با استفاده از این روش ترکیبی پیشنهاد شدند، نسبت‌به سبدهای با بازدهی به‌طور متوسط، بازدهی بالاتری دارند. اما باید اشاره کرد که در اینجا تعداد شرکت‌های هر سبد سهام بسیار محدود بوده و لذا می‌توان با استفاده از این روش ترکیبی مهم‌ترین متغیرهای موجود در مدل‌بندی بازده سهام را جست‌وجو و آن‌ها را در پرتفوی‌های با تعداد سهام بیشتر موردپژوهش قرار داد.

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

۹ پیشنهادهای کاربردی

ازآنجایی‌که بورس اوراق بهادار به‌عنوان یکی از مهم‌ترین مؤلفه‌های اقتصادی در هر کشور است و از سویی دیگر، عملکرد شرکت‌های فعال در کشور نیز مستقیماً درعملکرد اقتصادی و رشد اقتصادی کشور تأثیر می‌گذارد، لذا این تحقیق می‌تواند به‌عنوان الگویی به‌منظور مطالعات بعدی باشد. لذا می‌توان پیشنهاد کرد:

- مسئولین اجرایی سازمان بورس اوراق بهادار با اجرای استراتژی‌ها و سازوکارهای مناسب به ارتقای سودآوری و افزایش بازدهی دارایی شرکت‌ها مبادرت ورزیده و از این طریق موجبات گسترش بازار سرمایه و تقویت رشد اقتصادی را فراهم کنند.
- مسئولین اجرایی با بهره‌گیری از روش تحلیل سلسله‌مراتبی اقدام به تشکیل پرتفوی در شرکت‌ها کرده و از این طریق زمینهٔ گسترش دامنهٔ انتخاب سهام شرکت‌ها را فراهم کنند.
- به‌جای ارزش دفتری از بهای تمام‌شده استفاده شود. زیرا بهای تمام‌شده نسبت‌به ارزش دفتری معیاری مناسب‌تر است، چرا که از ارزش دفتری استهلاک کسر می‌شود، درحالی‌که رقم بهای تمام‌شده نسبت‌به ارزش دفتری رقمی بالاست. پس اگر در محاسبهٔ متغیرها بهای تمام‌شده در نظر گرفته شود، می‌تواند سهام‌دار را در تحلیل نرخ بازده سهام موردنظرشان (نزدیک به واقعیت) یاری رساند.

منابع و مآخذ

اصل، حسن و معصوم‌زاده، نسیم (۱۳۸۹). «پیش‌بینی احتمال تغییر قیمت سهام با استفاده از رگرسیون لجستیک در بورس اوراق بهادار تهران»، *تحقیقات حسابداری*، بهار ۱۳۸۹.

ایزدی نیا، ناصر؛ ابراهیمی، محمد؛ و حاجیان‌نژاد، امین (۱۳۹۳). «مقایسهٔ مدل اصلی سه عاملی فاما و فرنچ با مدل اصلی چهار عاملی کارهارت در تبیین بازده سهام شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران»، مقالهٔ ۳، دورهٔ ۲، شمارهٔ ۳، صص. ۱۷ـ۲۸.

ایمانی، م. (۱۳۹۰). «مقایسه عملکرد مدل D-CAPM با مدل سه عاملی فاما وفرنچ در پیش‌بینی بازده موردانتظار»، پایان‌نامهٔ کارشناسی ارشد، دانشگاه آزاد واحد تهران مرکز.

باقرزاده، سعید (۱۳۸۴). «عوامل مؤثر در بازده سهام در بورس اوراق بهادار تهران»، *تحقیقات مالی*، شمارهٔ ۱۹، صص. ۲۵ـ۶۴.

خانی، ابراهیم و ابراهیم‌زاده، آسو (۱۳۹۰). «آزمون مدل شرطی چندعاملی CAPM در بورس اوراق بهادار تهران»، *فصلنامهٔ بورس اوراق بهادار*، ش. ۱۶، زمستان.

راعی، رضا و بستان آرا، مهدی (۱۳۹۲). *مقایسهٔ کارایی مدل‌های رگرسیون با رویکرد تحلیل مؤلفه‌های اصلی (PCA) و شبکه‌های عصبی مصنوعی در پیش‌بینی بازده غیرعادی*، دانشگاه تهران

[DOR: 20.1001.1.26453355.1399.13.45.2.5]

راعی، رضا و چاوشی، کاظم (۱۳۸۲). «پیش‌بینی بازده سهام در بورس اوراق بهادار تهران: مدل شبکه‌های عصبی مصنوعی و مدل چندعاملی»، *تحقیقات مالی*، شمارهٔ ۱۵.
راعی، رضا و فلاح‌پور، سعید (۱۳۸۷). «کاربرد ماشین بردار پشتیبان در پیش‌بینی درماندگی مالی شرکت‌ها با استفاده از نسبت‌های مالی»، *بررسی‌های حسابداری و حسابرسی*، دانشکده مدیریت دانشگاه تهران، دورهٔ ۱۵، شمارهٔ ۵۳، صص. ۱۷ـ۳۴.
عادل، آذر و افسر، امیر (۱۳۸۵). «مقایسهٔ روش‌های کلاسیک و هوش مصنوعی در پیش‌بینی شاخص قیمت سهام و طراحی مدل ترکیبی»، *فصلنامه مدرس علوم انسانی*، شمارهٔ ۴.
عسکری راد، حسین (۱۳۹۲). «اثر عامل مومنتوم در توان توضیحی الگوی سه‌عاملی فاما و فرنچ: شواهدی از بورس اوراق بهادار تهران»، *فصلنامه دانش حسابداری*، ۴ (۱۲).
فشاری، مجید و مظاهری‌فر، پوریا (۱۳۹۵). «مقایسهٔ الگوریتم‌های پیش‌بینی و بهینه‌سازی در بورس اوراق بهادار تهران»، *دوفصلنامه سیاست‌گذاری پیشرفت اقتصادی*، دورهٔ ۴، شمارهٔ ۲ (۱۱)، صص. ۷۶-۵۹.
فلاح‌پور، سعید؛ گلارضی، غلامحسین؛ و فتوره‌چیان، ناصر (۱۳۹۲). «پیش‌بینی روند حرکتی قیمت سهام با استفاده از ماشین بردار پشتیبان بر پایه الگوریتم ژنتیک در بورس اوراق بهادار تهران»، *تحقیقات مالی*، دانشکده مدیریت دانشگاه تهران، دورهٔ ۱۵، شمارهٔ ۲، صص. ۲۶۹ـ۲۸۸.
قائمی، محمدحسین و طوسی، سعید (۱۳۸۵). «بررسی عوامل مؤثر در بازده سهام عادی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران»، *پیام مدیریت*، شماره‌های ۱۷ و ۱۸، صص. ۱۵۹ـ ۱۷۵.
لشکری، زهرا و نظام‌الاسلامی، حسین (۱۳۹۵). «عوامل مؤثر در رتبه‌بندی سهام صنایع مادر پذیرفته‌شده در بورس اوراق بهادار تهران با استفاده از F Topsis»، پایان‌نامهٔ کارشناسی ارشد، دانشگاه آزاد اسلامی واحد تهران مرکزی.
مجتهد زاده، ویدا و طارمی، مریم (۱۳۸۴). «آزمون مدل سه‌عاملی فاما و فرنچ در بورس اوراق بهادار تهران جهت پیش‌بینی بازده سهام»، *پیام مدیریت*، زمستان ۸۴ و بهار ۸۵، صص. ۱۰۹ـ۱۳۲.
مجتهدزاده، ویدا و ولی‌زاده، اعظم (۱۳۹۳). «تبیین مدلی برای پیش‌بینی بازده سهام»، پایان‌نامهٔ دکترا؛ دانشگاه الزهرا.
محمدی، شاپور و سعیدی، حسن (۱۳۹۱). «پیش‌بینی نوسانات بازده بازار با استفاده از مدل‌های ترکیبی گارچ- شبکهٔ عصبی»، *فصلنامه بورس اوراق بهادار*، شمارهٔ ۱۶.
نمازی، محمد و کیامهر، محمدمهدی (۱۳۸۶). «پیش‌بینی بازده روزانهٔ سهام شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران با استفاده از شبکه‌های عصبی مصنوعی»، *تحقیقات مالی*، دورهٔ ۹، شمارهٔ ۲۴.
نمازی، محمد و محمدتبار کاسگری، حسن (۱۳۸۶). «به‌کارگیری مدل چندعاملی برای توضیح بازده شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران»، *مجله علوم اجتماعی و انسانی دانشگاه شیراز*، ۲۶ (۱).

Brieman, L., J. Friedman, R. Olshen, and C. Stone (1984). *Classification and Regression Trees,* New York: Chapman & Hall,

Dash, Rajashree. Sidharth Samal, Rasmita Dash, Rasmita Rautray (2019). "An integrated TOPSIS crow search based classifier ensemble: In application to stock index price movement prediction", *Applied Soft Computing Journal*, S1568-4946(19)30565-4.

Zhu, M. (2012). "Jackknife for Bias Reduction in Predictive Regressions." *Forthcoming Journal of Financial Econometrics*.

Zhu, M., D. Philpotts, R. Sparks and M.J. Stevenson (2011). "A hybrid Approach to Combining Cart and Logistic Regression for Stock Ranking." *Journal of Portfolio Management*, Fall 38(1), 100 - 109.

Zhu, M., D. Philpotts and M.J. Stevenson (2012). "Classification and Regression Trees and Their Use in Financial Modeling." Book Chapter. Forthcoming in F.J. Fabozzi (Eds.) *Encyclopedia of Financial Models*. Wiley.

Zhou, Liang. Yain-Whar Si, Hamido Fujita (2017). Predicting the listing statuses of Chinese-listed companies using decision trees combined with an improved filter feature selection method; Knowledge-Based Systems 0 0 0 (2017) 1–9.

Sorensen, E., Mezrich, J. and Miller, K. (1998). *A new technique for tactical asset allocation. In F. Fabozzi (ed.), Active Equity Portfolio Management*, New Hope, Pennsylvania, pp. 195–202. 11.

Cochrane, J. (2011). "Presidential address: Discount rates." *Journal of Finance,* 66 (4), 1047– 1108. 1, 11.

Fama, E. F. and French, K. R. (1993). "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33 (1), 3–56. 9, 127.

[DOR: 20.1001.1.26453355.1399.13.45.2.5]