Journal of Financial Management Strategy
Vol. 3, No. 10
Autumn 2015

Alzahra University- Faculty of Social Sciences and Economics
Received: 2015/09/05
Accepted: 2015/11/16

Prediction of Stock Return Using Financial Ratios: A Decision Tree Approach

Ali Mohammad Alimohammadi[1]
Mohammad Hussein Abbasimehr[2]
Ahmad Javaheri[3]

Abstract

The purpose of this paper is to develop a model for prediction of present and prospect stock return using financial ratios. For this purpose, decision tree method was used. In this approach, a set of logical conditions in a hierarchical algorithmic model have been used for prediction or recognition of an event. Hereupon in this research, 70 percent of data were used to produce models in four popular decision tree algorithms (CHAID, ECHAID, QUEST and CRT) and the results of the tests were compared in 30 percent of residual of data with some of performance measures like accuracy, sensitivity and specificity. Information of 317 companies accepted in Tehran Stock Exchange was used in this study. The results indicated that ECHAID and CRT algorithms performed best in the prediction of present and CHAID algorithm in that of future. It was also shown that the models were better in that of present compared to that of future. The abilities of the models, however, were not significant in both cases. Accordingly, the hypothesis of the study was rejected.

Keywords: Prediction of stock return, indicate of contemporary return, financial ratios, decision tree algorithm.
JEL: G11,C4

---

1. MSc. Accounting- Azad university Qazvin branch.
2. MSc. Accounting- Shahid beheshti university (Corresponding Author).
3. MSc. Financial Management- Olome Eghtesadi faculty.

راهبرد مدیریت مالی
سال سوم، شماره یازدهم
زمستان ۱۳۹۴
صص ۱۵۱-۱۲۹

دانشگاه الزهرا (س)
دانشکده علوم اجتماعی و اقتصادی
تاریخ دریافت: ۱۳۹۴/۰۶/۱۴
تاریخ تصویب: ۱۳۹۴/۰۸/۲۵

پیش‌بینی بازده سهام شرکت‌ها با استفاده از نسبت‌های مالی تحت رویکرد درخت تصمیم

علی‌محمد علیمحمدی[۱]، محمدحسین عباسی مهر[۲] و احمد جواهری[۳]

چکیده

هدف از پژوهش حاضر این است که با استفاده از نسبت‌های مالی، به مدلی بر پایه نسبت‌های مالی برای پیش‌بینی بازده معاصر و آتی شرکت‌ها دست بیابیم. در این پژوهش به‌منظور بررسی توانایی نسبت‌های مالی در تبیین بازده معاصر و پیش‌بینی بازده آتی سهام، از روش درخت تصمیم استفاده شده است. در این روش مجموعه‌ای از شرط‌های منطقی به‌صورت یک الگوریتم با ساختار درختی برای پیش‌بینی و تبیین یک پیامد به‌کار می‌رود. ازاین‌رو مدل‌های حاصل از چهار الگوریتم درخت تصمیم (شامل CHAID، ECHAID، QUEST و CRT) با استفاده از ۷۰ درصد داده‌های پژوهش شکل گرفته و نتایج حاصل از آزمون آن‌ها در ۳۰ درصد باقیمانده داده‌ها به‌وسیله معیارهایی نظیر صحت، دقت و جداول درهم‌ریختگی مقایسه شده است. از اطلاعات ۳۱۷ شرکت پذیرفته شده بورس اوراق بهادار تهران طی سال‌های ۱۳۸۱ تا ۱۳۹۲ در این پژوهش استفاده شده است. نتایج به‌دست آمده نشان می‌دهد که الگوریتم‌های CRT و ECHAID در تبیین بازده معاصر و الگوریتم CHAID در پیش‌بینی بازده آتی بهترین عملکرد را دارند. همچنین قدرت مدل‌ها در تبیین بازده معاصر بیشتر از پیش‌بینی بازده آتی است. هرچند که در هر دو حالت توانایی مدل‌ها از نظر آماری قابل اتکا نبوده و ازاین‌رو فرضیه برقراری ارتباط تبیینی بین نسبت‌های مالی طرح شده در این پژوهش و تغییرات بازده معاصر و آتی سهام رد می‌شود.

واژه‌های کلیدی: پیش‌بینی بازده آتی، تبیین بازده معاصر، نسبت‌های مالی، درخت تصمیم.
طبقه‌بندی موضوعی: G11, G44

---

۱. کارشناس ارشد حسابداری- دانشگاه آزاد اسلامی واحد قزوین، Alimohamadi8821@gmail.com
۲. کارشناس ارشد حسابداری- دانشگاه شهید بهشتی، mhabbasimehr@gmail.com
۳. کارشناس ارشد مدیریت مالی- دانشکده علوم اقتصادی A.javaheri1366@gmail.com

مقدمه

از مهم‌ترین منابع و سرمایه‌های پولی موجود در جامعه، نقدینگی خرد و کلان پراکنده شده در سطح آن جامعه است. بدون شک جذب و هدایت صحیح این نقدینگی مستلزم تحلیل مناسب فرصت‌های سرمایه‌گذاری در بازارهای پولی و مالی جامعه است. شناخت فرصت‌های مناسب سرمایه‌گذاری در بازار اوراق بهادار نیازمند تحلیل دقیق صورت‌های مالی شرکت‌ها و پیش‌بینی سود و بازده آتی شرکت‌هاست (لگزیان و همکاران، ۱۳۹۰). نسبت‌های مالی ابزار سنتی ولی همچنان قوی و متداول برای تجزیه‌وتحلیل صورت‌های مالی است که از دیرباز مورد علاقه استفاده‌کنندگانی نظیر تحلیل‌گران مالی، اعتباردهندگان، سرمایه‌گذاران و مدیران مالی بوده است. در این تحلیل‌ها، به‌جای به‌کارگیری ارقام کلی صورت‌های مالی، از نسبت‌های ترکیبی این ارقام به‌منظور دستیابی به تصویر واضحی از شرکت استفاده می‌شود. همچنین با کمک روش‌هایی نظیر تحلیل افقی و عمودی و مقایسه شرکت با سایر رقبا، فهم دقیق‌تری از وضعیت مالی شرکت برای استفاده‌کنندگان فراهم می‌شود (دلن و همکاران،۲۰۱۳). اگرچه این نسبت‌ها به‌سادگی محاسبه می‌شوند، لیکن تفسیر آن‌ها اغلب مشکل و بحث‌برانگیز است؛ به ویژه زمانی که دو یا چند نسبت، نتایج متضادی داشته باشند.

از سویی با افزایش مبادلات تجاری، اقتصادی و پیشرفت‌های فنّاوری اطلاعات، حجم زیادی از داده‌های مالی به‌سرعت انباشته‌شده و هزینه نگهداری و پردازش اطلاعات نیز به طرز شگفت‌انگیزی کاهش داشته است. به‌طوری‌که هزینه نگهداری اطلاعات از یک دلار بابت هر مگابایت به چند پنی کاهش‌یافته و سرعت پردازش هر ۲۴ ماه یک‌بار دو برابر شده است(شیفرت،۲۰۰۴). این موضوع به‌نوبه خود مشکلاتی را برای استفاده بهینه و کارا از این داده‌ها به‌وجودآورده است. ازاین‌رو به‌منظور استفاده بهینه و مؤثر از داده‌های مالی در تصمیم‌گیری‌ها، از فن‌های داده‌کاوی استفاده می‌شود. داده‌کاوی فرآیندی است که با نگرشی نو به مسئله استخراج اطلاعات از داده‌های حجیم پرداخته و به کمک مجموعه‌ای از روش‌های آماری و مدل‌سازی می‌تواند الگوها و روابط پنهان موجود درداده‌ها را در کمترین زمان ممکن و با دقتی بالا تشخیص دهد. هدف داده‌کاوی، جستجو برای یافتن اطلاعاتی باارزش برای تصمیم‌گیری در یک پایگاه داده است. این فرآیند می‌تواند برای پیش‌بینی رویه‌ها و رفتارهای آتی در بازار مالی مورد استفاده قرارگیرد (باقرپور ولاشانی و همکاران،۱۳۹۱). با توجه به مطالب مذکور، پرسش‌های پژوهش به صورت زیر شکل می‌گیرد:

۱. آیا می‌توان با استفاده از نسبت‌های مالی به مدلی برای تبیین تغییرات بازده سهام شرکت‌ها دست یافت؟

۲. آیا می‌توان با استفاده از نسبت‌های مالی به مدلی برای پیش‌بینی تغییرات بازده سهام شرکت‌ها دست یافت؟

مروری بر مبانی نظری و پیشینه پژوهش

هر پیش‌بینی مستلزم شناسایی متغیرهای مؤثر بر متغیر هدف و انتخاب مدلی برای به‌کارگیری آن متغیرهاست. اطلاعات صورت‌های مالی مهم‌ترین و قابل‌اتکاترین اطلاعات در ارزشیابی شرکت محسوب می‌شود. یکی از انواع روش‌های تجزیه‌وتحلیل اطلاعات حسابداری، تجزیه‌وتحلیل نسبت‌های مالی است که از اواخر قرن نوزدهم میلادی در بین تحلیل‌گران رایج شده است و بنیان آن به طور کامل بر صورت‌های مالی شامل ترازنامه، صورت سود و زیان و صورت گردش وجوه نقد قرار دارد (مهرانی و همکاران،۱۳۸۳).

هوریگان (۱۹۶۵) معتقد بود که نسبت‌های مالی به‌عنوان محصول نظام حسابداری می‌بایست به شاخص منحصربه‌فردی برای ارزیابی عملکرد شرکت‌ها در ایالات‌متحده تبدیل شود. همچنین از نسبت‌های مالی می توان به‌عنوان داده‌های خام برای طراحی مدل‌های پیش‌بینی بازده آتی نیز استفاده کرد (آلتمن، ۱۹۶۸؛ بیور، ۱۹۶۶). از همین رو، پژوهشگران با استفاده از مجموعه‌های نسبت‌های مالی و روش‌های گوناگون آماری، تلاش زیادی برای دستیابی به مدلی‌هایی برای پیش‌بینی سود و بازده آتی و احتمال ورشکستگی شرکت‌ها نموده‌اند. گرچه به علت سهولت محاسبه و سادگی در فهم، استفاده از نسبت‌های مالی به‌شدت رایج است، اما در مورد اثربخشی نتایج این معیارها، نگرانی‌ها و محدودیت‌های زیر وجود دارد:

۱. هر گروه از نسبت‌های مالی تنها یک قسمت از فعالیت واحد تجاری را ارزیابی می‌کند. در نتیجه این روش تجزیه‌وتحلیل برای تعیین کارآمدی یا ناکارآمدی کل شرکت، به تنهایی از توانایی کمتری برخوردار است (هو و همکاران،۲۰۰۵).

۲. تعداد نسبت‌های مالی قابل استخراج از صورت‌های مالی زیاد بوده و هیچ فهرست پذیرفته شده در خصوص نوع، روش‌های محاسبه و تعداد نسبت‌های مالی وجود ندارد(هو و همکاران،۲۰۰۵). به‌عنوان مثال، گمبولا وکتز (۱۹۸۳) برای یافتن مدلی از نسبت‌های مالی در صنایع خرده‌فروشی و تولیدی از ۵۸ نسبت استفاده نمودند. درحالی‌که وایلهو و وو (۲۰۰۶) از ۵۹ نسبت، سینکا و همکاران (۲۰۰۵) از ۱۶ نسبت، اویار و اوکوموس (۲۰۱۰) از ۱۵ نسبت و کاراکا و سیگدم (۲۰۱۲) از ۲۴ نسبت استفاده نمودند. با این حال، بیشتر مقاله‌ها

و پژوهش‌های چاپ شده در مجله‌های معتبر، بین ۲۰ تا ۳۰ مورد از متداول‌ترین نسبت‌ها را مورد استفاده قرار داده‌اند

۳. ساختار نسبت‌های مالی در صنایع مختلف متفاوت است. بنابراین نمی‌توان از آن‌ها به‌صورت عام در سطح بازار استفاده نمود. شواهدی نشان می‌دهد که ساختار الگوهای نسبت‌های مالی در شرکت‌های خرده‌فروشی و تولیدکنندگان عمده، تفاوت دارند (گمبولا وکتز، ۱۹۸۳).

۴. پژوهش‌ها نشان داده است که این نسبت‌ها به‌شدت تحت تأثیر عوامل برون‌زا هستند. برای مثال سینکا و همکاران (۲۰۰۵) ثابت نمودند که اندازه شرکت و کشوری که شرکت در آن واقع‌شده است، بر ساختار نسبت‌های مالی تأثیرگذار هستند.

۵. ماهیت نسبت‌های مالی به علت استفاده از متغیرهای مشابه در محاسبات، همپوشانی بالایی با یکدیگر دارد. این موضوع سبب شده است تا در بسیاری از پژوهش‌ها، به علت عدم استفاده از روش‌های آماری مناسب، نتایج پژوهش تحت تأثیر خطاهایی نظیر غیر نرمال بودن نمونه، ناهمسانی واریانس و هم خطی قرار گیرد.

۶. اگرچه بسیاری از پژوهش‌های انجام شده، اثربخشی به کارگیری نسبت‌های مالی در پیش‌بینی عملکرد و تغییرات ارزش شرکت‌ها را مشخص نموده‌اند، ولی بیشتر آنها اغلب قادر به تشخیص و توضیح نقش هر یک از این نسبت‌های مالی نبوده‌اند (دیلن و همکاران،۲۰۱۳).

از این رو می‌بایست نسبت‌های مالی با روشی به کار رود که بتواند این محدودیت‌ها را برطرف نماید. در این پژوهش از مشهورترین روش‌های درخت تصمیم برای این موضوع استفاده می‌کنیم.

در بررسی‌های نخست، به طور عمده پژوهشگران از فن‌های آماری سنتی (برای مثال روش طراحی عوامل، آنووا، رگرسیون خطی و غیره) استفاده می‌کردند. بعدها از روش‌های پیشرفته‌تری در این پژوهش‌ها از روش‌های طراحی شده برای فرضیات غیرخطی و غیر نرمال استفاده شد. به طور مثال، آلتمن (۱۹۶۸) از تحلیل مبین چندگانه استفاده کرده است که داده‌ها را در شرایط غیر نرمال، دارای ناهمسانی واریانس و با در نظر گرفتن خودهمبستگی متغیرها مورد تجزیه‌وتحلیل قرار می‌دهد. پس از این دوره، پژوهش‌ها بر استفاده از فن‌های گوناگون داده‌کاوی مانند درخت تصمیم و شبکه عصبی نیز تمرکز نمودند.

چانگ(۲۰۰۶)برای رتبه‌بندی عملکرد بانک‌های تجاری در کشور تایوان از رویکرد سیستم خاکستری استفاده کرد و با استفاده از نسبت‌های مالی به‌عنوان شاخص‌های ارزیابی، رتبه بانک‌ها را تعیین نمود. نتایج پژوهش وی نشان می‌دهد که رویکرد سیستم خاکستری بهتر از روش‌های آماری رایج نظیر تحلیل رگرسیون، تحلیل عاملی و سایر روش‌های آماری چند متغیره می‌تواند عملکرد را ارزیابی کند. زیرا فاقد محدودیت‌های این روش‌ها یعنی وجود حجم زیادی از داده‌هاست.

سان و هوی(۲۰۰۶) تلاش کردند تا مشکلات مالی شرکت‌های پذیرفته شده در بورس چین را با استفاده از درخت تصمیم و بر پایه الگوریتم ژنتیک پیش‌بینی کنند. دقت مدل حاصل از پژوهش آن‌ها بالا و در حد ۹۴ درصد بود.

ونگ و همکاران(۲۰۰۹) برای پیش‌بینی بازده‌های سهام، با استفاده از پنجاه نسبت مالی از مدل درخت تصمیم بهره بردند. آن‌ها در پژوهش خود به مقایسه مدل‌های حاصل از چندین روش اجرای درخت تصمیم پرداختند. پژوهش آن‌ها نشان داد که تکنیک درخت تصمیم کیسه‌ای از سایر روش‌ها عملکرد بهتری دارد.

یو و ونجوآن (۲۰۱۰) از درخت تصمیم استفاده نمودند تا مؤثرترین نسبت مالی بر رشد سود شرکت‌های پذیرفته شده در بورس را تعیین کنند. آنها از مدل C5 به عنوان یکی از فن‌های درخت تصمیم استفاده نمودند. نتایج آزمون مدل حاصل از درخت تصمیم در این پژوهش نشان داد که مدل دقت بالایی (حدود ۹۴٪) در پیش‌بینی رشد سود شرکت‌ها دارد.

دیلن و همکاران (۲۰۱۳) در مقاله‌ای با عنوان ارزیابی عملکرد واحد تجاری با استفاده از نسبت‌های مالی و روش درخت تصمیم، با بهره‌گیری از چهار الگوریتم این فن، قدرت پیش‌بینی کنندگی نسبت‌های مالی را در مورد شاخص‌های ارزیابی عملکرد (بازده حقوق صاحبان سهام و بازده دارایی‌ها) مورد بررسی قراردادند. نتایج به‌دست آمده در مورد بازده حقوق صاحبان سهام نشان داد که مهم‌ترین نسبت‌های مالی به ترتیب سود قبل از مالیات به حقوق صاحبان سهام، حاشیه سود خالص، نسبت اهرم مالی و نسبت رشد فروش هستند. در خصوص بازده دارایی‌ها نیز مهم‌ترین نسبت‌ها به ترتیب سود قبل از مالیات به حقوق صاحبان سهام، حاشیه سود خالص، نسبت بدهی و نسبت گردش دارایی هستند.

در پژوهش‌های داخلی، پیرامون پیش‌بینی قیمت سهام به طور عمده از روش‌های سنتی مثل رگرسیون خطی استفاده شده و از بین روش‌های نوین نیز بیشتر تمرکز بر مدل‌های عصبی ،عصبی –

فازی و الگوریتم ژنتیک، فن‌هایی چون سیستم خاکستری و یا تحلیل پوششی داده‌ها بوده است و از مدل درخت تصمیم در پیش‌بینی بازده استفاده نشده است. معدود پژوهش‌های انجام شده در این حوزه طی سال‌های اخیر به شرح زیر است:

زیبانژاد و همکاران(۱۳۹۰) با به‌کارگیری فن خوشه‌بندی و رگرسیون درخت تصمیم (CRT) ورشکستگی مالی را با استفاده از نسبت‌های مالی در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران طی سال‌های ۱۳۸۱ تا ۱۳۸۵پیش‌بینی نمودند و در این زمینه مؤثرترین متغیرها را مشخص کردند. دقت مدل استخراجی آن‌ها در پیش‌بینی ورشکستگی حدود ۹۵٪ بود.

آهنگری (۱۳۹۰) در مقاله‌ای با استفاده از تکنیک درخت تصمیم به پیش‌بینی شرکت‌های ورشکسته و غیر ورشکسته در بازه زمانی ۱۳۸۴ تا ۱۳۸۸ پرداخت. وی در این پژوهش از تکنیک‌های C5، CHAID، QUEST، CRT استفاده نمود. نتایج پژوهش وی نشان داد که دقت مدل حاصل از تکنیک CRT بیشتر از سایر تکنیک ها و حدود ۹۴٪ می‌باشد.

باقری و همکاران (۱۳۹۰) با استفاده از تکنیک شبکه عصبی و درخت تصمیم به بررسی عوامل مؤثر بر به هنگام بودن گزارش‌های مالی پرداختند. نتایج پژوهش آن‌ها نشان داد که میانگین دقت شبکه عصبی در پیش‌بینی بهنگام بودن بالاتر از درخت تصمیم است.

رستگار مقدم مؤدب (۱۳۹۰) در پایان‌نامه‌ای با استفاده از شبکه عصبی مصنوعی و درخت تصمیم به پیش‌بینی مدیریت سود در شرکت‌های بورس اوراق بهادار تهران در بازه سال‌های ۱۳۸۰ تا ۱۳۸۸ پرداخت. نتایج پژوهش وی نشان داد که توان پیش‌بینی مدل غیرخطی شبکه عصبی مصنوعی و درخت تصمیم نسبت به مدل رگرسیون خطی چندگانه بیشتر است.

حجازی و همکاران (۱۳۹۱) نیز در پژوهشی مشابه به مقایسه توانایی مدل‌های حاصل از رگرسیون خطی و تکنیک درخت تصمیم در پیش‌بینی مدیریت سود پرداختند. بازه زمانی مورد مطالعه آن‌ها ۱۳۸۵ تا ۱۳۸۸ و تعداد اعضای نمونه ۶۶۰ فصل- شرکت بود. پژوهش آن‌ها نیز تأیید کرد که قدرت مدل‌های مبتنی بر تکنیک درخت تصمیم در پیش‌بینی مدیریت سود بیشتر از مدل‌های مبتنی بر رگرسیون خطی است.

در این پژوهش، برای اولین بار از تکنیک های درخت تصمیم برای پیش‌بینی بازده سهام شرکت‌ها در بازار بورس اوراق بهادار تهران استفاده شده است. همچنین در این پژوهش ، از یک مجموعه داده مالی گسترده و غنی از شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران استفاده نموده‌ایم.

فرضیه پژوهش

با توجه به مطالب یاد شده در بخش های قبلی، فرضیه پژوهش به صورت زیر شکل می‌گیرد:

۱. با استفاده از نسبت‌های مالی می‌توان تغییرات بازده سهام شرکت‌ها در دوره معاصر را تبیین نمود.

۲. با استفاده از نسبت‌های مالی می‌توان تغییرات بازده سهام شرکت‌ها در دوره آتی را پیش‌بینی نمود.

روش‌شناسی پژوهش

در جدول ۱ به طور خلاصه، مراحل پژوهش نشان داده شده است. به این ترتیب که ابتدا ۳۱ نسبت مالی به شرح جدول ۲ محاسبه شده و مورد پردازش قرار گرفته‌اند. این نسبت‌ها مبتنی بر اطلاعات صورت‌های مالی است که داده‌های آن از نرم‌افزار رهاورد نوین استخراج شده است. این متغیرها به‌عنوان ورودی (متغیر مستقل) و بازده تعدیل شده سهام معاصر و آتی به‌عنوان متغیر خروجی (متغیر وابسته) تعیین شدند. بازده معاصر عبارت است از بازده تعدیل شده سهام شرکت در بازه یک‌ساله منتهی به ارائه صورت‌های مالی و منظور از بازده آتی، بازدهی سهام شرکت طی سه ماه بعد از انتشار صورت‌های مالی است. این دو متغیر به نوعی نشان دهنده ارزش عملکرد مالی فعلی و آتی شرکت‌ها هستند. اطلاعات بازده شرکت‌ها نیز به کمک شاخص قیمت سهام شرکت‌ها (منتشر شده توسط مرکز پژوهش، توسعه و مطالعات اسلامی سازمان بورس) محاسبه شده است. برای دستیابی به نتایج واقعی و قابل‌استفاده ابتدا الگوریتم‌های درخت تصمیم با استفاده از ۷۰ درصد از نمونه که داده‌های آموزشی (Training set) نامیده می‌شوند، برای بازده معاصر و بازده آتی برازش شدند و سپس برای تعیین عملکرد مدل‌های حاصله در دنیای واقعی، دیگر داده‌ها را (که به آنها داده‌های آزمایشی (Testing set) می‌گوییم) در مدل‌های به‌دست آمده آزمون شدند و از این طریق نتایج عملکرد مدل‌های حاصل از هر روش مورد بررسی قرار گرفت. مدل‌های درخت تصمیم با استفاده از نرم‌افزار اس پی اس اس طبق چهار الگوریتم E-CHAID ، CHAID، QUEST، CRT استخراج و در نهایت به منظور انتخاب دقیق‌ترین مدل ، قدرت مدل‌ها در پیش‌بینی و تبیین بازده مورد مقایسه قرار گرفته است.

جدول ۱- مراحل پژوهش

داده‌های خام
استخراج از پایگاه داده نرم‌افزار رهاورد نوین و مرکز پژوهش، توسعه و مطالعات اسلامی سازمان بورس
↓
پردازش اولیه
آماده سازی داده‌ها،اعمال تعدیلات لازم،حذف داده‌های اضافی و نامرتبط و قالب بندی داده‌ها
↓
فازبندی مجموعه داده‌ها
در نظر گرفتن ۷۰ ٪ داده‌ها به عنوان داده‌های آموزشی و ۳۰٪ داده‌ها به عنوان داده‌های آزمایشی
↓
ساخت مدل با استفاده از داده‌های آموزشی (Training Data)
تحلیل درخت تصمیم- مدل‌های پیش‌بینی کننده : CHAID – RT - C5&C – QUEST
↓
آزمودن مدل با استفاده از داده‌های آزمایشی (Testing Data)
استفاده از ماتریس درهم‌ریختگی برای مقایسه عملکرد مدل‌ها در قالب جدول

جدول ۲- نسبت‌های مالی مورد استفاده در پژوهش

نسبت مالی	مدل محاسباتی
نسبت‌های نقد شوندگی	
نسبت آنی	بدهی‌های جاری ÷ (موجودی کالا – دارایی‌های جاری)
نسبت جاری (نقد شوندگی)	بدهی‌های جاری ÷ دارایی‌های جاری
نسبت نقد	بدهی‌های جاری ÷ وجه نقد و معادل‌های آن
نسبت‌های گردش یا بهره‌وری دارایی‌ها	
گردش حساب‌های دریافتنی	حساب‌های دریافتنی ÷ فروش
گردش موجودی کالا	موجودی کالا ÷ بهای تمام شده کالای فروش رفته
گردش سرمایه در گردش خالص	(بدهی‌های جاری – دارایی‌های جاری) ÷ فروش
گردش دارایی‌ها	کل دارایی‌ها ÷ فروش
گردش مالکانه	حقوق مالکانه ÷ فروش
گردش دارایی‌های ثابت	دارایی‌های ثابت ÷ فروش
گردش دارایی‌های غیرجاری	دارایی‌های غیرجاری ÷ فروش
گردش دارایی‌های جاری	دارایی‌های جاری ÷ فروش

ادامه جدول ۲

نسبت‌های سودآوری	
حاشیه سود ناخالص	فروش ÷ سود ناخالص
حاشیه سود(EBITDA به فروش)	سود قبل از هزینه‌های مالی و مالیات و استهلاک به فروش
حاشیه سود خالص	فروش ÷ سود خالص
سود قبل از مالیات به حقوق صاحبان سهام	حقوق صاحبان سهام ÷ سود قبل از مالیات
نرخ بازده حقوق صاحبان سهام	حقوق صاحبان سهام ÷ سود خالص
نرخ بازده دارایی‌ها	دارایی‌ها ÷ سود خالص
هزینه‌های عملیاتی به فروش خالص	فروش خالص ÷ هزینه‌های عملیاتی
نسبت‌های رشد	
نرخ رشد دارایی‌ها	دارایی‌های سال قبل ÷ (دارایی‌های سال قبل – دارایی‌های سال جاری)
نرخ رشد سود خالص	سود خالص سال قبل ÷ (سود خالص سال قبل – سود خالص سال جاری)
نرخ رشد فروش	فروش سال قبل ÷ (فروش سال قبل – فروش سال جاری)
نسبت‌های ساختار دارایی‌ها	
دارایی جاری به کل دارایی‌ها	کل دارایی‌ها ÷ دارایی‌های جاری
موجودی کالا به دارایی جاری	دارایی‌های جاری ÷ موجودی کالا
وجه نقد و معادل‌هایش به دارایی‌های جاری	دارایی های جاری ÷ وجه نقد و معادل‌های آن
دارایی‌های غیرجاری به کل دارایی‌ها	کل دارایی‌ها ÷ دارایی‌های غیر جاری
نسبت‌های توان پرداخت بدهی‌ها	
بدهی مالی کوتاه‌مدت به کل بدهی‌ها	کل بدهی‌ها ÷ بدهی مالی کوتاه‌مدت
بدهی جاری به کل بدهی‌ها	کل بدهی‌ها ÷ بدهی جاری
نسبت پوشش هزینه‌های مالی	هزینه‌های مالی ÷ سود قبل از هزینه‌های مالی و مالیات
بدهی کل به حقوق صاحبان سهام	حقوق صاحبان سهام ÷ بدهی کل
نسبت بدهی (اهرمی)	دارایی کل ÷ بدهی کل
کل بدهی مالی به کل بدهی‌ها	کل بدهی‌ها ÷ کل بدهی مالی

داده‌ها و نمونه

هدف مقاله، شناسایی و استفاده از یک مجموعه داده بزرگ و غنی بوده است. نمونه اولیه این پژوهش شامل کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران به‌جز شرکت‌های فعال در صنعت بانکی، واسطه‌های مالی، شرکت‌های سرمایه‌گذاری و شرکت‌های فعال در صنعت بیمه و بازنشستگی از سال ۸۱ تا پایان سال ۹۲ بوده است. درمجموع از اطلاعات ۳۱۷ شرکت برای محاسبه متغیرهای پژوهش استفاده شد که با توجه به محدودیت تعریف شده در نمونه گیری (منفی نبودن حقوق صاحبان سهام) در نهایت نسبت‌های مالی برای ۲۷۲۰ سال شرکت محاسبه شد. اما به دلیل عدم امکان محاسبه بازده برای تمامی شرکت‌های مذکور، تعداد نمونه برای بررسی بازده معاصر به ۲۵۰۲ و برای بررسی بازده آتی به ۲۲۵۶ سال شرکت تقلیل یافته است. در جدول ۳، مختصات کلی شرکت‌های موجود در نمونه به‌وسیله آمار توصیف شده است. بدیهی است که به دلیل تعدد نسبت‌های مالی محاسبه شده امکان نشان دادن آمار توصیفی تمام آن‌ها مقدور نیست.

جدول ۳- آمار توصیفی

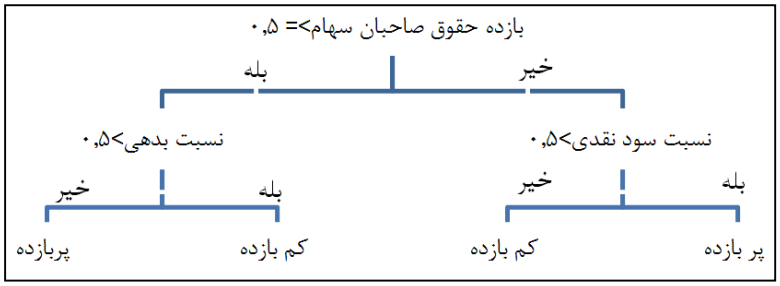
شرح	جمع ترازنامه	جمع حقوق صاحبان سهام	سود و زیان خالص	جمع درآمد	نسبت جاری	بازده	
						سال گذشته	سه ماهه آتی
واحد	میلیارد ریال	میلیارد ریال	میلیارد ریال	میلیارد ریال	-	درصد	درصد
میانگین	۱،۹۸۲	۷٦٥	۲۸۲	۱،٦۸۱	۱,۲۳۳۲	٤٥،۰۹	۱۲،۰۲
انحراف معیار	۷،۸۱۳	۳،۲۷۸	۱،۳۱٦	۸،٦۰۷	۰،٦٤٤۰	۱۱٤،٦۷	۳٤،۱٦
چارک اول	۱٦۱	٤٥	۸	۱۲۳	۰،۸۸٦۸	-۸،٤٦	-٤،٤۸
میانه	۳۷٤	۱۱٤	۳۳	۲۷۷	۱،۱۲۷۱	۱٤،۷۰	۳،۰۱
چارک سوم	۹٥٥	۳۲۱	۱۱٤	۷۱٥	۱،۳۹٤۷	٦۱،٥۱	۱۹،٥۹
بیشترین	۱۱۳،۷٦۸	٥٤،۰۸٥	۲۳،٥۰٤	۱۹۲،٥۲٥	۹،۹۹۳۱	۱۹۷٥،٦۱	۳٦۸،٤۹
کمترین	۱٥	۰،۱۸٤	-۷،۲۰٤	۰،٥۳۹	۰،۲۰۲۲	-۸۸،٦۲	-٦٤،۹٤
تعداد شرکت	۳۱۷	۳۱۷	۳۱۷	۳۱۷	۳۱۷	۳۱٤	۲۹۹

الگوریتم‌های درخت تصمیم

درخت تصمیم یکی از روش‌های ناپارامتریک رده‌بندی کردن است که با توجه به نوع متغیر وابسته به دو دسته رده‌بندی درخت برای متغیر رسته‌ای و رگرسیون درختی برای متغیر پیوسته تقسیم می‌شود. رده‌بندی درختی در راستای روش‌هایی نظیر تحلیل ممیزی(تابع تشخیص) و رگرسیون لجستیک است. در این روش مجموعه‌ای از شرط‌های منطقی به‌صورت یک الگوریتم با ساختار درختی برای رده‌بندی یا پیش‌بینی یک پیامد به کار می‌رود(بهنام پور و همکاران، ۱۳۹۲).

در درخت تصمیم‌گیری، تعدادی پرسش وجود دارد و با مشخص شدن پاسخ هر پرسش، پرسشی دیگر طرح می‌شود. اگر پرسش‌ها درست و سنجیده مطرح شوند، تعداد کمی از پرسش‌ها برای پیش‌بینی دسته رکورد جدید کافی خواهد بود. عملکرد درخت تصمیم به این صورت است که یک گره ریشه در بالای آن کشیده شده و برگ‌های آن در پایین قرار دارند. هر نمونه‌ای که در گره ریشه وارد می‌شود، در یک آزمون قرار می‌گیرد تا روشن شود که این رکورد به کدام یک از گره‌های بعدی ارسال شود. در جدول ۴، نمایی از یک درخت تصمیم ساده برای انتخاب شرکت پربازده دیده می‌شود:

جدول ٤- نمایی از یک درخت تصمیم



برخلاف دسته‌بندی کننده‌های تک مرحله‌ای رایج که هر نمونه روی تمام دسته‌ها امتحان می‌شود، دریک درخت دسته‌بندی کننده ، هر قسمت از نمونه (با توجه به مشخصه‌های آن) روی زیرمجموعه‌های خاصی از دسته‌ها امتحان شده و محاسبات غیر لازم حذف می‌شوند. در دسته‌بندی کننده‌های تک مرحله‌ای فقط از زیر مجموعه‌ای صفات برای روش بین دسته ها استفاده می‌شود که به طور معمول با یک معیار بهینه سراسری انتخاب می‌شود. در دسته‌بندی کننده درخت

انعطاف‌پذیری انتخاب زیرمجموعه‌های مختلفی از صفات در گروه‌های داخلی مختلف درخت وجود دارد، که زیرمجموعه انتخاب شده به شکل بهینه بین دسته‌های این گروه را تفکیک می‌کند. این انعطاف‌پذیری ممکن است کارایی مدل نهایی را نسبت به دسته‌بندی کننده‌های تک مرحله‌ای افزایش دهد (دستگیر و شفیعی سردشت،۱۳۹۰). علاوه بر موارد فوق، این الگوریتم‌ها مزایایی به شرح زیر دارند:

۱. برای تقریب توابع گسسته بکار می‌رود و نسبت به نویز داده‌های ورودی مقاوم است.
۲. برای داده‌های با حجم بالا کاراست. ازاین‌رو در داده‌کاوی استفاده می‌شود.
۳. می‌توان درخت را به‌صورت قوانین علت و معلولی و قابل‌فهم برای استفاده‌کننده نمایش داد.
۴. امکان ترکیب عطفی و فصلی فرضیه‌ها را می‌دهد.
۵. در مواردی نیز قابل‌استفاده است که نمونه‌های آموزشی فاقد همه ویژگی‌ها هستند(بردواج و پال،۲۰۱۲؛ دیلن و همکاران،۲۰۱۳)

به خاطر همین ویژگی‌ها از نظر بسیاری از سازمان‌های حرفه‌ای، این فن ها به‌عنوان روش‌های برتر شناخته شده است. برای مثال، انجمن حسابداران رسمی آمریکا، فن‌های داده‌کاوی از جمله درخت تصمیم را به‌عنوان یکی از ده تکنیک فناورانه برتر برای آینده معرفی کرده است. همچنین انجمن حسابرسان داخلی آمریکا نیز این فن را در فهرست یکی از چهار اولویت پژوهشی خود گنجانده است (کیرکدس و همکاران،۲۰۰۴).

اگرچه الگوریتم‌های درخت تصمیم بسیار زیادی وجود دارند ولی در این پژوهش از الگوریتم‌های مشهور و پرکاربردی نظیر CHAID ، EXHAUSTIVE CHAID ، CRT و QEUST استفاده می‌شود.

CHAID[1]: این روش یک فن آماری بسیار مؤثر است که توسط کاس (۱۹۸۰) ابداع شده است. این روش یک فن درخت تصمیم مبتنی بر شناسایی قوی‌ترین روابط متقابل بین متغیرهای مستقل و وابسته است و به این منظور از مقدار احتمال آماره کای دو مربوط به آزمون استقلال جداول توافقی استفاده می‌کند. در این روش از بین متغیرهای موجود، متغیر دارای مقدار P-Value کوچک‌تر در مرحله نخست برای تقسیمات روی یک گره در نظر گرفته می‌شود. وجه تمایز

1. Chi-squared Automatic Interaction Detector

این الگوریتم با سایر فن‌های درخت تصمیم آن است که CHAID می‌تواند در هر سطح از درخت بیش از دو طبقه ایجاد نماید. در نتیجه این الگوریتم، یک روش درختی دودویی (باینری) محسوب نمی‌شود. به همین خاطردرخت ایجاد شده در این روش وسیع‌تر از سایر روش‌هاست. خروجی این الگوریتم بسیار عینی و تفسیر آن ساده است. زیرا در این روش، به طور پیش‌فرض از انشعاب‌های چندگانه استفاده می‌شود(اس پی اس اس،۲۰۰۷). از سوی دیگر ضعف این روش در عدم توانایی آن در ایجاد بهینه‌ترین تقسیمات بر اساس متغیرهای موجود است (بهنام‌پور و همکاران، ۱۳۹۲).

E-CHAID[1]: این فن روش بهبودیافته CHAID است که در آن تمامی انشعاب‌های ممکن برای هر متغیر مستقل آزمون می‌شود. از این‌رو ساختار درخت ایجاد شده در این روش دقیق‌تر از روش CHAID است(اس پی اس اس،۲۰۰۷).

CRT[2]: درخت‌های رده‌بندی و رگرسیون توسط بریمن و همکاران (۱۹۸۴) کشف شدند. این الگو، یک الگوریتم درخت تصمیم دودویی (باینری) باقابلیت پردازش و طبقه‌بندی متوالی متغیرهای مستقل است. منظور از پردازش متوالی داده‌ها این است که داده‌ها بر اساس یک متغیر به دو زیرمجموعه تقسیم می‌شوند تا میزان همگنی آنها در هر زیرمجموعه نسبت به زیرمجموعه قبلی افزایش یابد. این دو زیرمجموعه نیز دوباره تجزیه می‌شوند و این امر تا زمانی ادامه می‌یابد که معیار همگنی یا سایر معیارهای متوقف کننده دستور توقف تجزیه را بدهند. هدف نهایی تجزیه، تعیین متغیر مناسب مربوط به آستانه مناسب برای حداکثر سازی همگنی زیرگروه‌های نمونه است. این الگوریتم، یک سلسله از درخت‌های هرس شده تودرتو ایجاد می‌کند که هر یک می‌تواند بهینه باشد. انتخاب درخت بهینه از طریق ارزیابی عملکرد هر یک از این درخت‌های هرس شده در مجموعه داده مستقل انجام می‌شود. این مکانیسم، متعادل‌سازی طبقه‌ها را به‌صورت خودکار و بهینه انجام می‌دهد(اس پی اس اس،۲۰۰۷). ضعف این روش اریب بودن آن در انتخاب متغیرهاست. علاوه بر این، در متغیرهای کیفی چند سطحی، نتایج حاصل از این روش گیج کننده خواهد بود. زیرا ممکن است چند سطح یک متغیر به یک گره تعلق بگیرد که این امر باعث می‌شود نتوان تفسیر ساده‌ای از نتایج ارائه نمود (بهنام پور و همکاران، ۱۳۹۲).

---

1. Exhaustive Chi-squared Automatic Interaction Detector
2. Classification and Regression Trees

QUEST[1]: درخت آماری مؤثر سریع و نااریب یک الگوریتم به نسبت جدید با تجزیه دودویی (باینری) برای رده‌بندی و داده‌کاوی است. این مدل، شباهت زیادی به الگوریتم CRT دارد. بااین‌حال تفاوت‌های کوچکی نیز دارد. ملاک انتخاب متغیرها در این روش برای متغیرهای کمی، مقدار احتمال مربوط به آماره F آزمون آنوا و برای متغیرهای کیفی احتمال آماره کای دو است ازاین‌رو درخت ایجاد شده در این روش نااریب بوده و ضمن داشتن دقت الگوریتم CRT با سرعت بیشتری محاسبه می‌شود (اس پی اس اس،۲۰۰۷).

تحلیل مقایسه‌ای درخت تصمیم

به‌منظور کنترل دقت درخت‌های تصمیم لازم است تا مقادیر متغیرهای وابسته از حالت پیوسته به دودویی (موفق یا ناموفق) تبدیل شود. ازاین‌رو مقدار میانه (معیار گرایش به مرکز) بازده به‌عنوان معیار تفکیک در نظر گرفته شده است. به این صورت که نمونه‌ای با بازدهی بالاتر از میانه در طبقه موفق و نمونه‌هایی با بازده عملکرد پایین‌تر از میانه، در طبقه ناموفق قرار داده شده‌اند. عملکرد مدل‌های دودویی (دوگروهی) اغلب با استفاده از یک ماتریس درهم‌ریختگی (جدول ۵) اندازه‌گیری می‌شود. ماتریس درهم‌ریختگی شامل اطلاعات ارزشمندی درباره رده‌بندی‌های واقعی و پیش‌بینی‌شده توسط مدل رده‌بندی است (کوهاوی و پروست، ۱۹۹۸). بر مبنای اطلاعات حاصل از این ماتریس شاخص‌های مقایسه مناسبی محاسبه می‌شود. شاخص‌هایی نظیر: نمودار بازدهی، صحت کلی (AC)، مساحت زیر نمودار (AUC)، معیار دقت، معیار تشخیص، معیار حساسیت و معیار F. این شاخص‌ها در **Error! Reference source not found.**جدول ۶ تعریف شده‌اند.

جدول ۵- ماتریس درهم‌ریختگی عملکرد مدل پیش‌بینی بازده

		پیش‌بینی‌شده	
		نامناسب	مناسب
واقعی	نامناسب	منفی صحیح(TN)	مثبت کاذب(FP)
	مناسب	منفی کاذب(FN)	مثبت صحیح(TP)

1. The Quick, Unbiased, Efficient Statistical Tree

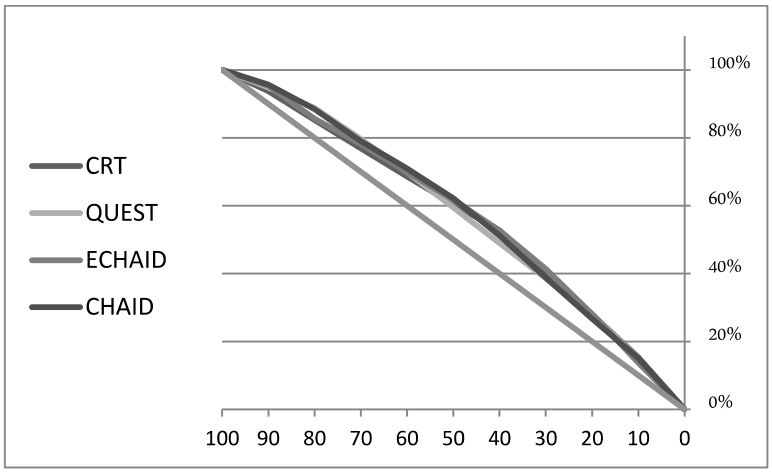
جدول ٦- تعریف معیارهای مقایسه مدل‌های درخت تصمیم

نام	نماد	تعریف	رابطه محاسباتی
صحت کلی	AC	درصد اطلاعاتی که به‌درستی توسط مدل پیش‌بینی شده‌اند	$\text{صحت} = \frac{TP+TN}{TP+TN+FP+FN}$
دقت	P	نسبت موارد پیش‌بینی‌شده مثبت صحیح به مجموع پیش‌بینی‌های مثبت (صحیح و کاذب)	$\text{دقت} = \frac{TP}{TP+FP}$
حساسیت	S	نرخ پیش‌بینی مثبت صحیح/ قدرت طبقه‌بندی مناسب رویدادهای خوب را نشان می‌دهد	$\text{حساسیت} = \frac{TP}{TP+FN}$
معیار	F	میانگین هارمونیک معیارهای عملکرد دقت و حساسیت	$F \text{ معیار} = ۲ \times \frac{\text{حساسیت} \times \text{دقت}}{\text{حساسیت} + \text{دقت}}$
تشخیص	-	نرخ موارد منفی صحیح (TN)/ قدرت طبقه‌بندی مناسب رویدادهای بد را نشان می‌دهد	$\text{نرخ منفی واقعی} = \frac{TN}{TN+FP}$
نرخ مثبت کاذب	-	نسبت تعداد مثبت کاذب به جمع موارد منفی	$\text{نرخ مثبت کاذب} = \frac{FP}{FP+TN}$
نسبت منفی کاذب	-	نسبت تعداد منفی کاذب به جمع موارد منفی	$\text{نرخ منفی کاذب} = \frac{FN}{FN+TP}$

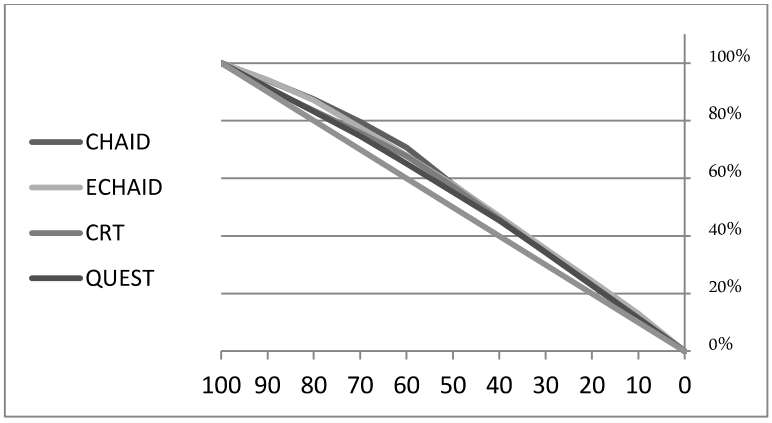
منحنی ROC نیز علاوه بر شاخص‌های فوق یکی از روش‌های مقایسه قدرت مدل‌های مختلف پیش‌بینی کننده است. چنانچه مرز بین احتمال وقوع رویدادهای بد و خوب بسیار متغیر باشد، معیارهای حساسیت و تشخیص نیز بسیار متغیر خواهد شد. منحنی ترکیب (۱-تشخیص، حساسیت) منحنی ROC نامیده می‌شود. مساحت زیر این منحنی یک شاخص ترکیبی است که نشان می‌دهد مدل با چه احتمالی موقعیت مثبت را نسبت به موقعیت منفی انتخاب می‌کند. بیشترین حد رقم این شاخص ۱ و کمترین حد آن برای انتخاب تصادفی ۰/۵ است. برای آنکه مدل مورد استفاده ازنظر آماری معنادار باشد باید این مقدار بالاتر از ۰/۷ باشد(اس پی اس اس، ۲۰۰۷).

نتایج تحلیل درخت تصمیم

در این پژوهش، ما از چهار الگوی درخت تصمیم CHAID، CRT، ECHAID و QUEST استفاده نمودیم تا بهترین مدل‌های رده‌بندی شرکت‌ها ازنظر بازده را شناسایی نماییم.نتایج تحلیل در دو مقطع مورد آزمایش قرار گرفت. در بخش نخست تحلیل، بازده معاصر به‌عنوان متغیر وابسته در نظر گرفته شد، درحالی‌که در بخش دوم، بازده آتی این نقش را ایفا نموده است.



نمودار ۱- ارزیابی داده‌های آزمایشی برای بازده معاصر



نمودار ۲-ارزیابی داده‌های آزمایشی برای بازده آتی

نمودار ۱ و نمودار ۲، عملکرد انواع روش‌های درخت تصمیم را با کمک نمودار بازدهی[1] نمایش می‌دهند. در این نمودار، محور عمودی میزان بازدهی و محور افقی توزیع داده‌ها به ترتیب میزان موفقیت را نشان می‌دهد. بازدهی در این نمودار به‌عنوان نسبت مجموع موارد موفق در هر چندک تعریف می‌شود و هر چه قوس رو به بالای نمودار یک روش بیشتر باشد (چندک‌های بیشتری، با موفقیت پیش‌بینی شده باشند) عملکرد روش مذکور بهتر بوده است. بازدهی در این نمودار به‌صورت زیر محاسبه می‌شود:

بازدهی =(تعداد کل موارد/تعداد موارد موفق در هر چندک)×۱۰۰

تراکم خطوط بازدهی روش‌های مختلف نشان دهنده نزدیکی عملکرد روش‌های مختلف است و نزدیکی آن‌ها به خط میانه که به‌صورت مستقیم از صفر تا صد کشیده شده است، نشان می‌دهد که قدرت تبیین و پیش‌بینی بازده در هر دو آزمون چندان زیاد نیست. ازاین‌رو لازم است تا برای بررسی بهتر عملکرد الگوریتم‌های درخت تصمیم، از معیارها و شاخص‌هایی بهره بجوییم که پیش‌تر توضیح داده شده است.

جدول ۷- نتایج مقایسه روش‌های مختلف پیش‌بینی بازده بر مبنای داده‌های آزمایشی برای بازده معاصر

انواع درخت تصمیم	TP	TN	FP	FN	صحت	دقت	حساسیت	معیار دقت F	تشخیص	نرخ مثبت کاذب	نرخ منفی کاذب	مساحت زیر منحنی
CHAID	۱۹۳	۲۴۳	۱۰۷	۱۷۰	۰/۶۱	۰/۶۴	۰/۵۳	۰/۵۸	۰/۶۹	۰/۳۱	۰/۴۷	۰/۶۳
ECHAID	۲۴۸	۱۹۶	۱۷۰	۱۰۳	۰/۶۲	۰/۵۹	۰/۷۱	۰/۶۴	۰/۵۴	۰/۴۶	۰/۲۹	۰/۶۳
CRT	۱۸۶	۲۷۹	۹۸	۱۸۳	۰/۶۲	۰/۶۵	۰/۵۰	۰/۵۷	۰/۷۴	۰/۲۶	۰/۵۰	۰/۶۴
QUEST	۱۸۶	۲۴۸	۹۸	۱۷۰	۰/۶۲	۰/۶۵	۰/۵۲	۰/۵۸	۰/۷۲	۰/۲۸	۰/۴۸	۰/۶۰
درخت بهینه					CRT	CRT	ECHAID	ECHAID	CRT	CRT	ECHAID	CRT

1. Gain chart

بررسی شاخص‌های مربوط به مدل‌های مربوط به بازده معاصر در جدول ۷ نشان می‌دهد که طبق معیارهای صحت کلی (AC)، دقت، تشخیص، نرخ مثبت کاذب و مساحت زیر نمودار، مدل CRT بهترین عملکرد را داشته است. برای مثال این مدل در ۶۳ درصد موارد، پیش‌بینی صحیحی از وضعیت بازده معاصر داشته است(معیار صحت) و یا ۶۵ درصد از بازده‌های مناسب بازار را شناسایی نموده است (معیار دقت) درحالی‌که از نظر سه معیار دیگر یعنی حساسیت، معیار F و نرخ منفی کاذب مدل حاصل از روش ECHAID عملکرد بهتری دارد. برای مثال این مدل در ۷۱ درصد از موارد طبقه‌بندی بهتری از رویدادهای خوب ارائه نموده است. بااین‌حال با توجه به مقادیر آماره مساحت زیر منحنی که جملگی کمتر از ۷ دهم هستند، می توان نتیجه گرفت که مدل‌های به‌دست آمده از نظر آماری چندان قابل‌اتکا نیستند. لیکن در این میان قدرت مدل CRT از سایر مدل‌ها بیشتر بوده است(جدول ۹).

ماتریس درهم‌ریختگی مدل‌ها برای بازده معاصر نیز نشان می‌دهد که کلیه مدل‌های درخت تصمیم، تنها در ۶۱ تا ۶۳ درصد موارد قادر به تبیین بازده معاصر هستند (سه ستون نخست جدول ۹). این جدول نیز همسو با شاخص‌های محاسبه‌شده، عملکرد روش CRT را – با توانایی پیش‌بینی ۶۲/۳ درصد از بازده‌ها– از سایر روش‌ها بهتر نشان می‌دهد

همچنین بررسی نتایج مدل‌های مربوط به بازده آتی در جدول ۸ نشان می‌دهد که به‌جز سه معیار تشخیص، نرخ مثبت کاذب و مساحت زیر منحنی، در دیگر معیارها روش CHAID از عملکرد بهتری برخوردار بوده است. درحالی‌که طبق معیارهای تشخیص و نرخ مثبت کاذب روش ECHAID و بر اساس معیار مساحت زیر نمودار، مدل CRT بهترین عملکرد را داشته است. بررسی مقادیر آماره مساحت زیر منحنی در این روش نیز نشان داد که مدل‌های به‌دست آمده از نظر آماری چندان قابل‌اتکا نیستند. لیکن در این میان قدرت مدل CRT از سایر مدل‌ها بیشتر بوده است که از این نظر هر دو آزمون مشابه هم بوده‌اند.

ماتریس‌های درهم‌ریختگی برای بازده آتی نیز نشان می‌دهد که توانایی پیش‌بینی کلیه مدل‌های درخت تصمیم، نسبت به آزمون قبلی کاهش‌یافته و تنها بین ۵۵ تا ۶۱ درصد از بازده‌ها را درست پیش‌بینی کرده‌اند (سه ستون دوم جدول ۹). این جدول نیز همسو با شاخص‌های محاسبه شده، عملکرد روش CHAID را – با توانایی پیش‌بینی ۶۰/۸ درصد از بازده‌ها– از سایر روش‌ها بهتر نشان می‌دهد.

جدول ۸- نتایج مقایسه روش‌های مختلف پیش‌بینی بازده بر مبنای داده‌های آزمایشی برای بازده آتی

انواع درخت تصمیم	TP	TN	FP	FN	صحت	دقت	حساسیت	معیار دقت F	تشخیص	نرخ مثبت کاذب	نرخ منفی کاذب	مساحت زیر منحنی
CHAID	۲٤۸	۱۷۰	۱۸۱	۸۸	۰/٦۱	۰/۵۸	۰/۷٤	۰/٦۵	۰/٤۸	۰/۵۲	۰/۲٦	۰/۵۹
ECHAID	۲۱٤	۱۹٦	۱۵۸	۱۳۸	۰/۵۸	۰/۵۸	۰/٦۱	۰/۵۹	۰/۵۵	۰/٤۵	۰/۳۹	۰/٦۰
CRT	۱۹۵	۱۷۰	۱٦٦	۱۰۵	۰/۵۷	۰/۵٤	۰/٦۵	۰/۵۹	۰/۵۱	۰/٤۹	۰/۳۵	۰/٦۱
QUEST	۲۵۸	۱۲۰	۲۰۹	۹۲	۰/۵٦	۰/۵۵	۰/۷٤	۰/٦۳	۰/۳٦	۰/٦٤	۰/۲٦	۰/۵٦
درخت بهینه					CHAID	CHAID	CHAID	CHAID	ECHAID	ECHAID	CHAID	CRT

جدول ۹- ماتریس‌های در هم‌ریختگی هر مدل درخت تصمیم بر مبنای داده‌های آزمایشی برای بازده معاصر(سه ستون نخست) و آتی (سه ستون دوم)

نوع روش	واقعی	تبیین بازده معاصر			پیش‌بینی بازده آتی		
		نامناسب (۰)	مناسب (۱)	درصد پیش‌بینی صحیح	نامناسب (۰)	مناسب (۱)	درصد پیش‌بینی صحیح
CHAID	نامناسب (۰)	۲٤۳	۱۰۷	٪٦۹/٤	۱۷۰	۱۸۱	٪٤۸/٤
	مناسب (۱)	۱۷۰	۱۹۳	٪۵۳/۲	۸۸	۲٤۸	٪۷۳/۸
	نسبت نوع پیش‌بینی به کل	٪۵۷/۹	٪٤۲/۱	٪٦۱/۲	٪۳۷/٦	٪٦۲/٤	٪٦۰/۸
E-CHAID	نامناسب (۰)	۱۹٦	۱۷۰	٪۵۳/٦	۱۹٦	۱۵۸	٪۵۵/٤
	مناسب (۱)	۱۰۳	۲٤۸	٪۷۰/۷	۱۳۸	۲۱٤	٪٦۰/۸
	نسبت نوع پیش‌بینی به کل	٪٤۱/۷	٪۵۸/۳	٪٦۱/۹	٪٤۷/۳	٪۵۲/۷	٪۵۸/۱
CRT	نامناسب (۰)	۲۷۹	۹۸	٪۷٤/۰	۱۷۰	۱٦٦	٪۵۰/٦
	مناسب (۱)	۱۸۳	۱۸٦	٪۵۰/٤	۱۰۵	۱۹۵	٪٦۵/۰
	نسبت نوع پیش‌بینی به کل	٪٦۱/۹	٪۳۸/۱	٪٦۲/۳	٪٤۳/۲	٪۵٦/۸	٪۵۷/٤
QUEST	نامناسب (۰)	۲٤۸	۹۸	٪۷۱/۷	۳۰۰	٤۹۹	٪۳۷/۵
	مناسب (۱)	۱۷۰	۱۸٦	٪۵۲/۲	۲۰۲	۵۷٦	٪۷٤/۰
	نسبت نوع پیش‌بینی به کل	٪۵۹/۵	٪٤۰/۵	٪٦۱/۸	٪۳۱/۸	٪٦۸/۲	٪۵۵/۵

یافته‌های پژوهش و نتیجه‌گیری

نتایج قسمت نخست پژوهش نشان داد که مدل‌های CRT و ECHAID در فراهم نمودن مدل تبیین بازده معاصر، موفق‌تر از سایر مدل‌ها بوده‌اند. درحالی‌که برای پیش‌بینی بازده دوره آتی مدل-های مبتنی بر روش CHAID مقادیر مناسب‌تری از معیارهای مورد اندازه‌گیری را کسب نموده است. لیکن در هر دو آزمون، روش CRT از نظر معیار مساحت زیر نمودار یکی از معتبرترین شاخص‌هاست؛ روش بهینه مدل‌سازی است. همچنین مقادیر معیارهای مقایسه‌ای نشان داد که قدرت پیش‌بینی کنندگی بازده آتی این مدل‌ها کمتر از توانایی تبیین بازده معاصر است. بااین‌وجود سطح آماره اخیر در کلیه روش‌ها زیر هفت دهم بوده که نشان می‌دهد استفاده از این فن درخت تصمیم برای تبیین بازده معاصر و پیش‌بینی بازده آتی مناسب نیست و از این رو سبب رد شدن آماری فرضیه‌های پژوهش می‌شود.

این پژوهش با فراهم نمودن شواهدی در خصوص عدم کارایی مدل‌های مبتنی بر الگوریتم درخت تصمیم نشان می‌دهد که اگرچه این روش در موضوعاتی نظیر پیش‌بینی ورشکستگی(زیبانژاد، ۱۳۹۰؛ آهنگری، ۱۳۹۰؛سان و هوی،۲۰۰۶)، مدیریت سود (رستگارمقدم مؤدب، ۱۳۹۰؛ حجازی و همکاران، ۱۳۹۱) و پیش‌بینی عملکرد مالی (دیلن و همکاران، ۲۰۱۳؛ یو و نجوآن،۲۰۱۰) سبب شکل‌گیری مدل‌های قوی در تبیین و پیش‌بینی متغیرهای وابسته شده است، لیکن توانایی این روش در دادن مدلی برای پیش‌بینی و تبیین بازده آتی و معاصر در بورس اوراق بهادار تهران تأیید نشد. این موضوع می‌تواند حاصل مواردی از قبیل: عدم تفکیک و طبقه‌بندی شرکت‌های حاضر در نمونه براساس صنعت یا معیارهای دیگری نظیر شاخص‌های شفافیت اطلاعات، عدم استفاده از سایر روش‌های تکنیک درخت تصمیم نظیر C5 و عدم تناسب مجموعه نسبت‌های مالی به کارگرفته شده در این پژوهش باشد.

براساس نتایج این پژوهش، برای پژوهش‌های آتی در این حوزه پیشنهاد می‌شود، نمونه مورد بررسی به روش‌های مختلف طبقه‌بندی شده و علاوه بر بکارگیری روش‌های دیگر تکنیک درخت تصمیم، از نسبت‌های مالی بازار نظیر قیمت به سود و... نیز استفاده شود. انتظار می‌رود اجرای این پیشنهادها سبب افزایش دقت مدل‌های نهایی شود.

منابع

- آهنگری، مهناز. (۱۳۹۰) بکارگیری الگوریتم درخت تصمیم جهت پیش‌بینی شرکت‌های ورشکسته و غیر ورشکسته پذیرفته شده در بورس اوراق بهادار تهران. اولین کنفرانس ملی دانش پژوهان کامپیوتر و فناوری اطلاعات، تبریز، دانشگاه تبریز.
- باقری، مصطفی؛ سلیمانی مارشک، مجتبی؛ مرادی، مهدی. (۱۳۹۰). بررسی عوامل مؤثر بر بهنگامی گزارش گری مالی با استفاده از تکنیک های شبکه عصبی مصنوعی و درخت تصمیم. *پژوهش های تجربی حسابداری*، ۳، ۱۵۹-۱۷۷.
- بهنام پور، ناصر؛ حاجی زاده، ابراهیم؛ سمنانی، شهریار؛ زایری. (۱۳۹۲). معرفی الگوریتم های مدل رده‌بندی درختی و کاربرد آن در تعیین عوامل مؤثر بر ابتلا به سرطان مری در استان گلستان. *جرجانی*، ۲، ۴۶-۵۵.
- حجازی، رضوان؛ محمدی، شاپور؛ اصلانی، زهرا؛ آقاجانی، مجید. (۱۳۹۱). پیش‌بینی مدیریت سود با استفاده از شبکه عصبی و درخت تصمیم در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. *بررسی های حسابداری و حسابرسی*، ۶۸، ۳۱-۴۶.
- دستگیر، محسن؛ شفیعی سردشت، مرتضی. (۱۳۹۰). فناوری داده‌کاوی؛ رویکردی نوین در حوزه مالی. *دانش حسابرسی*، ۵، ۶-۲۷.
- رستگار مقدم مؤدب، هیوا. (۱۳۹۰) *پیش‌بینی مدیریت سود با استفاده از شبکه های عصبی مصنوعی و درخت تصمیم برای شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران*. رساله کارشناسی ارشد: دانشکده علوم اقتصادی.
- زیبانژاد، یاسر؛ فروغی، داود؛ مونادجمعی، اکبر. (۱۳۹۰). اجرای فن درخت تصمیم برای پیش‌بینی ورشکستگی. *دانش کامپیوتر و مهندسی اتوماسیون*، ۴، ۱۶۵-۱۶۹.
- لگزیان، محمد؛ بقایی، جواد؛ همایونی راد، محمدحسین. (۱۳۹۰). بررسی تأثیر نسبت‌های مالی بر پیش‌بینی سود شرکت و بازده سهام. *دو فصلنامه اقتصاد پولی، مالی*، ۱، ۱۰۲-۱۱۵.
- مهرانی، ساسان؛ مهرانی،کاوه؛کرمی، غلامرضا. (۱۳۸۳). استفاده از اطلاعات تاریخی مالی و غیرمالی جهت تفکیک شرکت‌های موفق از ناموفق. *بررسی‌های حسابداری و حسابرسی*، ۳۸، ۷۷-۹۲.

- باقرپور ولاشانی، محمد علی؛ ساعی، محمد جواد؛ مشکانی، علی؛ باقری، مصطفی. (۱۳۹۱) پیش‌بینی گزارش حسابرس مستقل در ایران، رویکرد داده کاوی. *تحقیقات حسابداری*، ۱۹، ۱۳۴-۱۵۰.
- Altman, Edward I. (1968). Financial ratios, discriminant analysis and the predication of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.
- Beaver, William H. (1966). Financial ratios as predictors of failure. Empirical Research in Accounting: Selected Studies. *Supplement to Journal of Accounting Research*, 4, 71–111.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*, New York: Chapman & Hall/CRC.
- Chang, C. p. (2006). Managing business attributes and performance for commercial banks. *Journal of American Academy of Business*, 9(7), 104-109.
- Cinca, C. S., Molinero, C. M., & Larraz, J. L. G. (2005). Country and size effects in financial ratios: A European perspective. *Global Finance Journal*, 16, 26-47.
- Delen, Dursan, Kuzey, Cemil, Uyar, Ali, (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert systems with application*, 40, 3970-3985.
- Gombola, Michael J., & Ketz, J. Edward. (1983). Financial ratio patterns in retail and manufacturing organizations. *Financial Management*, 12(2), 45-56.
- Horrigan, James O. (1965). Some empirical bases of financial ratio analysis. *The Accounting Review*, 40(3), 558-568.
- Karaca, S. S., & Çigdem, R. (2012). The effects of the 2008 world crisis to Turkish ˘certain sectors: The case of food, main metal, stone and soil and textile industries. *International Research Journal of Finance and Economics* , (88), 59-68.
- Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 119-127.
- Kirkos S. and Manolopoulos Y. (2004). Data Mining in Finance and Accounting: A Review of Current Research Trends, Proceedings of the 1st International Conference on Enterprise Systems and Accounting (ICES Acc), Thessaloniki,Greece.
- Kohavi; Provost. (1998). Glossary of terms, editorial for the special issue on applications of machine learning and the knowledge discovery process. *Machine Learning*, 30(2-3), 271-274.
- Seifert ,Jeffrey W. (2004). Data Mining: An Overview. Available on line At: Www.Fas.Org/Irp/Crs/Rl31798.Pdf.
- SPSS. (2007). *Clementine12 User Manual*, Chicago, IL

- Uyar, A., & Okumus, E. (2010). Finansal Oranlar Aracılıg ˘ıyla Küresel Ekonomik Krizin Üretim S. irketlerine Etkilerinin Analizi: IMKB'de Bir Uygulama, *Muhasebe ve Finansman Dergisi*. vol. 46, April 2010, pp. 146–156.
- Wang, H., Jiang, Y., & Wang, H. (2009). Stock return prediction based on bagging- decision tree, In Proceedings of 2009 IEEE international conference on grey systems and intelligent services, November 10-12, Nanjing, China.
- Wu, D., Liang, L., Huang, Z., Li, S. X.,(2005). Aggregated Ratio Analysis in DEA, *International Journal of Information Technology & Decision Making* ,4(3),369-384.
- Yu, G., & Wenjuan, G. (2010). Decision tree method in financial analysis of listed logistics companie, In 2010 International conference on intelligent computation technology and automation, Changsha-China.