

چشم‌انداز مدیریت مالی

شماره ۱۴ - تابستان ۱۳۹۵

صص ۸۷-۱۰۳

ارزیابی مدل‌های درخت تصمیم در پیش‌بینی عملکرد مالی شرکت‌های بورس اوراق بهادار تهران

ایمان رئیسی وانانی*، قاسم بولو**، شهره زرکش***

چکیده

نسبت‌های مالی همواره یکی از منابع قوی در ارزیابی عملکرد مالی شرکت‌های بورس اوراق بهادار تهران است. یکی از روش‌های پیش‌بینی عملکرد استفاده از الگوریتم‌های داده‌کاوی است. در این پژوهش، چهار مدل درخت تصمیم به‌منظور ارزیابی عملکرد، پیاده‌سازی و مدل‌ها با معیارهای ارزیابی مقایسه شدند. بدین منظور نمونه‌ای متشکل ۲۱ نسبت در ۵۳۴ شرکت پذیرفته‌شده در بورس اوراق بهادار تهران در فاصله بین سال‌های ۱۳۹۰ تا ۱۳۹۳ به‌عنوان متغیرهای مستقل و دو نسبت بازده دارایی‌ها و بازده حقوق صاحبان سهام به‌عنوان 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 به‌دست‌آمده از صحت بالاتری برخوردار است و در بین چهار درخت تصمیم سی‌فایو از بهترین شاخصه‌های ارزیابی برخوردار بود.

کلیدواژه‌ها: ارزیابی عملکرد مالی؛ درخت تصمیم؛ داده‌کاوی.

تاریخ دریافت مقاله: ۱۳۹۵/۰۷/۱۰، تاریخ پذیرش مقاله: ۱۳۹۵/۰۹/۲۶

* استادیار مدیریت صنعتی، دانشگاه علامه طباطبائی.

** دانشیار مدیریت و حسابداری، دانشگاه علامه طباطبائی.

*** دانشجوی کارشناسی ارشد مدیریت مالی، دانشگاه علامه طباطبائی (نویسنده مسئول).

E-mail: Shohreh.zarkesh@gmail.com

۱. مقدمه

در عصر حاضر حوزه مدیریت مالی نیز همگام با حوزه فناوری اطلاعات در حال پیشرفت است و با وجود تحولات شگرف دانش وجود نظام ارزیابی موضوعی اجتناب‌ناپذیر است. شرکت‌ها برای حضور در عرصه بازار سرمایه باید بتوانند همواره اصل بهبود مستمر را سرلوحه فعالیت‌های خود قرار دهند. ارزیابی عملکرد یکی از مباحث مهم در توانمندسازی و قابلیت پاسخگویی در چارچوب اصول و مفاهیم مدیریت است. در سال‌های اخیر تغییرات چشمگیری در زمینه‌های مدیریت مالی، سیستم‌های اطلاعاتی و آمار صورت گرفته است. افزایش تصاعدی در قدرت محاسباتی باعث شده است تا سیستم‌های اطلاعاتی هوشمند بتوانند در زمان اندک حجم زیادی از داده‌ها را پردازش کنند. تکنیک‌های داده‌کاوی ابزارهای شناخته‌شده‌ای برای تجزیه و تحلیل داده‌ها هستند. این تکنیک‌ها با استفاده از روش‌های مختلف، الگوهای موجود در مجموعه داده‌های بسیار بزرگ را استخراج می‌کنند؛ لذا پژوهش حاضر با به‌کارگیری روش‌های درخت تصمیم در داده‌کاوی به دنبال بهترین الگوریتم برای پیش‌بینی عملکرد مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران است. داده‌کاوی در حقیقت کشف ساختارهای جالب توجه، غیرمنتظره و با ارزش از درون مجموعه وسیعی از داده‌ها است [۱۰].

انعطاف‌پذیری و قابلیت فهم بودن یکی از برتری‌های روش درخت تصمیم است [۹]. در این تحقیق با توجه به اهمیت مشخص بودن الگوها و قواعد و سهولت در تحلیل این روش انتخاب شده است. به‌طور کلی درخت تصمیم، روشی برای نمایش مجموعه‌ای از قوانین اگر آنگاه است که منجر به دستیابی به هدف می‌شود. از آنجاکه اطلاعات مالی به‌صورت خام قابل تحلیل نیستند و استفاده از آن‌ها باعث آشفتگی می‌شود باید در ابتدا دانش نهفته در داده‌ها استخراج شود.

۲. مبانی نظری و پیشینه پژوهش

درخت تصمیم یکی از ساده‌ترین اما قدرتمندترین روش‌های تجزیه و تحلیل متغیرهای چندگانه [۴] و یکی از روش‌های داده‌کاوی است که اغلب به‌منظور طبقه‌بندی و پیش‌بینی به‌کار می‌رود. در دنیای تجارت داده‌های مالی به‌عنوان سرمایه راهبردی مطرح هستند. در دنیای کنونی روزانه حجم وسیعی از داده‌ها تولید می‌شود. یکی از نیازهای مهم تحلیل این داده‌های حجیم است. امروزه داده‌کاوی در درجه اول توسط شرکت‌هایی استفاده می‌شود که تمرکز اصلی آن‌ها روی مصرف‌کننده، مباحث مالی، حوزه ارتباطات و بازاریابی است. در این پژوهش با استفاده از چهار درخت تصمیم به تبیین بهترین الگوریتم در پیش‌بینی عملکرد مالی پرداخته شده است و مدل‌های به‌دست‌آمده با سه روش مورد ارزیابی قرار گرفته شده است.

کاشف تعاملات خودکار کای دو^۱ یک الگوریتم یادگیری با ناظر است. این روش یک روش آماری بسیار موثر برای تقسیم‌بندی، یا رشد درختی است و توسط کاس در سال ۱۹۸۰ توسعه یافته است؛ این الگوریتم از تست آماری برای تصمیم در جداسازی و از جداکننده‌های چند جانبه برای ساخت درخت استفاده می‌کند؛ همچنین دارای یک مکانیسم توقف است که تصمیم می‌گیرد یک زیر درخت چه هنگام کامل شده است [۲۲].

الگوریتم سی فایو^۲ نیز یک یادگیرنده با ناظر است که توسط راس کوینلن در سال ۱۹۳۸ به منظور ساخت یک درخت تصمیم با استفاده از مفهوم شاخص افزایش اطلاعات ساخته شد. این الگوریتم با جداسازی داده بر اساس بیشترین شاخص افزایش اطلاعات کار می‌کند. این روند تا جایی پیش می‌رود که هیچ زیرمجموعه‌ای را نتوان جداسازی کرد.

الگوریتم آماری موثر عاری از تعصب و سریع^۳ در سال ۱۹۹۷ توسط لو و شی برای متغیرهای پاسخ اسمی طراحی شد. درخت رده‌بندی حاصل از این الگوریتم نظیر مدل کارت دارای تقسیمات دودویی بوده و ملاک تصمیم برای انتخاب متغیرها با استفاده از مقدار p مربوط به آماره F آزمون تجزیه و تحلیل واریانس برای متغیرهای کمی و p آماره کای-دو مربوط به جداول توافقی برای متغیرهای کیفی صورت می‌پذیرد. این الگوریتم با توجه به اینکه از مقدار p برای تصمیم‌گیری استفاده می‌نماید، موجب تشکیل درختی ناریب از متغیرها می‌شود. این الگوریتم ضمن حفظ دقت برآورد در مدل کارت، از سرعت بالاتری در معرفی یک درخت رده‌بندی نسبت به آن برخوردار است.

درخت خوشه‌بندی و رگرسیون^۴ توسط بریمن، فردمن، اولشن و استون (۱۹۸۴) آغاز شد و یک الگوریتم درخت تصمیم دودویی مستعد فرآیند متوالی یا پیش‌بینی‌کننده مطلق با متغیرهای هدف است و به‌طور بازگشتی کار می‌کند [۳]. داده‌ها به دو زیرمجموعه گروه‌بندی می‌شوند تا سوابق را همگن‌تر از زیرمجموعه قبلی بکند؛ سپس دو زیرمجموعه دوباره تجزیه می‌شود تا زمانی که معیار همگن بودن یا دیگر معیارهای توقف رضایت‌بخش باشد. بخش پیش‌بینی‌کننده مشابه ممکن است چندین بار در درخت استفاده شود. هدف غایی تجزیه تعیین متغیر صحیح وابسته به آستانه درست برای ماکزیمم کردن همگن بودن گروه‌های زیرمجموعه نمونه است. علاوه بر آن این الگوریتم مقدارهای از قلم افتاده را با استفاده از تجزیه جانشین کنترل می‌کند تا بهترین استفاده از داده را داشته باشد.

شناسایی نسبت‌های مالی موثر بر عملکرد شرکت‌ها همواره مورد توجه سرمایه‌گذاران و ذی‌نفعان بوده است. تمرکز تحقیقات گذشته بیشتر بر اساس تکنیک‌های آماری مانند تحلیل

1. CHAID
2. C5.0
3. Quest
4. C&R

عاملی، رگرسیون خطی و تجزیه و تحلیلی واریانس بوده است. بر اساس ابزارهای فناوری اطلاعات نوین و بانک‌های اطلاعاتی تحقیقات اخیر به سمت مباحث داده کاوی رفته است.

در سال ۲۰۱۳ دولت‌آبادی و همکاران با استفاده از مدل درخت تصمیم و رگرسیون لاجستیک به پیش‌بینی ورشکستگی مالی شرکت‌های بورسی در بازه زمانی ۲۰۰۶ تا ۲۰۱۱ پرداختند. با توجه به نتایج درخت تصمیم با دقت ۷۷/۶ درصد برای پیش‌بینی وضعیت شرکت و رگرسیون لاجستیک با دقت ۸۴/۵ درصد قابلیت پیش‌بینی داشتند [۶].

دلن و همکاران (۲۰۱۳) در پژوهشی با عنوان «ارزیابی عملکرد واحد تجاری با استفاده از نسبت‌های مالی با رویکرد درخت تصمیم» به مقایسه الگوریتم‌های درخت تصمیم پرداخته و با استفاده از نسبت‌های مالی، عملکرد شرکت‌های بورس استانبول را بررسی کردند. نتایج نشان می‌دهد نسبت سود قبل از مالیات به حقوق صاحبان سهام و حاشیه سود خالص دو متغیر اساسی در پیش‌بینی عملکرد شرکت‌ها هستند [۵].

کوینسگی و ازگلباس (۲۰۱۲) با استفاده از درخت تصمیم CHAID شرکت‌های کوچک و متوسط را در سال ۲۰۰۷ از لحاظ بحران مالی بررسی کردند و با استفاده از ۳۱ نسبت مالی یک سیستم هشدار به هنگام بحران طراحی کردند [۱۸].

یو و ونجوان (۲۰۱۰) برای بررسی اینکه کدام‌یک از نسبت‌های مالی قوی‌ترین تاثیر را بر رشد سود شرکت‌های تدارکاتی دارند، از روش C5.0 که یکی از روش‌های درخت تصمیم است استفاده کردند و به بررسی تحلیل‌های مالی از جمله تجزیه و تحلیل سوددهی پرداختند و نتایج نشان داد روش C5.0 دارای دقت ۹۵٪ است [۲۶].

ونگ 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 [۲۷].

کادویا و همکاران (۲۰۰۸) در پژوهشی استراتژی‌های متضاد سرمایه‌گذاری را با استفاده از تکنیک تجزیه و تحلیل پوششی داده‌ها مورد ارزیابی قرار دادند و ضمن معرفی معیارهای سنتی ارزیابی سرمایه‌گذاری مانند ارزش دفتری به قیمت بازار، استفاده از دو شاخص قدیم و تحلیل پوششی داده‌ها را برای ارزیابی مفید می‌دانند [۱۵].

جوسته (۲۰۰۶) در پژوهشی با عنوان «نسبت‌های جریان نقدی به‌عنوان یک معیار برای ارزیابی عملکرد مالی در شرکت‌های آفریقایی» شرکت‌های موجود در صنایع مختلف در کشورهای در حال توسعه آفریقایی را با شرکت‌های مشابه در آمریکا از لحاظ نسبت‌های نقدینگی مورد بررسی قرار داد. یافته‌های این پژوهش نشان داد که نسبت کفایت جریان نقدی حاکی از

آن است که صنایع آفریقای جنوبی برخلاف صنایع آمریکایی دارای وجه نقد کافی برای پرداخت تعهدات ضروری، هستند [۱۴].

لئون (۲۰۰۴) طی تحقیق خود نشان داد نسبت‌های مالی توانایی پیش‌بینی بازده سهام را دارند. با استفاده از رگرسیون و حداقل مربعات بازده نقدی در سال‌های ۱۹۴۶-۲۰۰۰ و ارزش دفتری به ارزش بازار و سود به قیمت در سال‌های ۱۹۶۳-۲۰۰۰ می‌توانستند به‌خوبی بازده را پیش‌بینی کنند [۱۹].

در تحقیقات اولیه اولسن و موسمن (۲۰۰۳) با استفاده از ۶۱ نسبت مالی، به پیش‌بینی بازده سهام ۲۳۵۲ شرکت کانادایی در بازه ۱۹۸۳ تا ۱۹۹۳ پرداختند. در این تحقیق شبکه عصبی با دو روش حداقل مربعات و تکنیک رگرسیون لاجستیک مقایسه شده است. نتایج نشان می‌دهد شبکه عصبی نسبت به دو روش دیگر تخمین بهتری را انجام می‌دهد [۲۴].

در ایران نیز طی سال‌های اخیر تحقیقاتی در حوزه مالی با استفاده از مباحث داده‌کاوی صورت گرفته است. صالحی و فرخی پیلرود (۱۳۹۳) در تحقیق خود به بررسی دقت پیش‌بینی مدیریت سود با استفاده از شبکه‌های عصبی و درخت تصمیم‌گیری و مقایسه آن‌ها با مدل‌های خطی پرداختند. بدین منظور ۹ متغیر تأثیرگذار بر مدیریت سود به‌عنوان متغیرهای مستقل و اقلام تعهدی اختیاری، به‌عنوان متغیر وابسته مورد استفاده قرار گرفته است. در این تحقیق دو صنعت کشاورزی و نساجی از سال ۱۳۸۵ تا سال ۱۳۹۰ مورد بررسی قرار گرفت. از روش رگرسیون کمترین مربعات برای مدل خطی و از شبکه عصبی پیشخور و درخت تصمیم‌گیری کارت برای بررسی از طریق روش‌های داده‌کاوی استفاده شد. نتایج نشان داد که روش شبکه عصبی و درخت تصمیم‌گیری در پیش‌بینی مدیریت سود نسبت به روش‌های خطی دقیق‌تر و دارای سطح خطای کمتری است. در رابطه با ارتباط بین متغیرهای وابسته با متغیر مستقل نیز می‌توان گفت، مدیریت سود با متغیرهای اقلام تعهدی اختیاری دوره قبل، اقلام تعهدی غیراختیاری دوره قبل یا آستانه عملکرد و حساسیت در پرداخت بابت عملکرد در روش‌های رگرسیون، شبکه عصبی، درخت تصمیم دارای بیشترین ارتباط است [۲۵].

کرمی و طلایی (۲۰۱۳) ارتباط بین نسبت‌های قیمت به سود، ارزش دفتری به ارزش بازار، نرخ سود نقدی و بازده سرمایه شرکت‌های بورس تهران را در بازه زمانی ۱۹۹۸ تا ۲۰۰۷ بررسی کرده است. نتایج نشان داد نسبت‌های ارزش دفتری به ارزش بازار و بازده سرمایه توانایی پیش‌بینی بازده سهام را دارند [۲۱].

حسینی و رشیدی (۱۳۹۲) با استفاده از درخت تصمیم کارت و رگرسیون لجستیک به پیش‌بینی ورشکستگی اقتصادی شرکت‌ها در بورس بین سال‌های ۱۳۷۸ تا ۱۳۸۹ پرداخته‌اند. هر دو مدل توانایی پیش‌بینی را داشته و رگرسیون لجستیک از عملکرد بهتری برخوردار بود [۱۳].

هاشمی و همکاران (۱۳۹۱) در تحقیق خود به مقایسه محتوای فزاینده اطلاعاتی نسبت‌های نقدی و تعهدی برای ارزیابی عملکرد مالی واحدهای تجاری با استفاده از مدل درخت تصمیم کارت و رگرسیون لجستیک پرداختند. نمونه آماری شامل ۱۹۹ شرکت از بین شرکت‌های پذیرفته‌شده در بورس بهادار تهران طی سال‌های ۱۳۸۳ تا ۱۳۸۸ است. یافته‌های پژوهش نشان می‌دهد نسبت‌های نقدی در پیش‌بینی عملکرد مالی شرکت‌های قوی و نسبت‌های تعدی در پیش‌بینی عملکرد مالی شرکت‌های ضعیف از محتوای اطلاعاتی بالاتری در هر دو مدل برخوردارند [۱۱].

حجازی و همکاران (۱۳۹۱) دقت پیش‌بینی مدیریت سود را با استفاده از شبکه‌های عصبی و درخت تصمیم‌گیری بررسی و همچنین با مدل‌های خطی مقایسه کردند. برای این منظور از یازده متغیر تاثیرگذار بر مدیریت سود به‌عنوان متغیرهای مستقل و اقلام تعهدی اختیاری به‌عنوان متغیر وابسته استفاده شده است. در این تحقیق تعداد ۵۵ شرکت از سال ۱۳۸۵ تا سال ۱۳۸۸ به‌صورت فصلی مورد بررسی قرار گرفته است. از روش رگرسیون پنلی جهت مدل خطی و از شبکه عصبی پیشخور تعمیم‌یافته و درخت تصمیم کارت جهت بررسی از طریق شبکه عصبی و درخت تصمیم استفاده شد. نتایج نشان داد که روش شبکه عصبی و درخت تصمیم در پیش‌بینی مدیریت سود نسبت به روش‌های خطی دقیق‌تر و دارای سطح خطای کمتری است [۱۲].

مهدوی و قربانی (۱۳۹۱) در پژوهشی با عنوان «بررسی مقایسه‌ای نقش شاخص‌های نوین و سنتی نقدینگی در ارزیابی عملکرد مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران»، ارتباط میان نسبت جاری، نسبت آنی (به‌عنوان شاخص سنتی نقدینگی) و همچنین شاخص فراگیر نقدینگی، دوره تبدیل وجه نقد، مانده نقدی خالص (به‌عنوان شاخص نوین نقدینگی) با بازده دارایی، بازده حقوق صاحبان سهام و کیوتوبین (به‌عنوان شاخص عملکرد) بررسی کردند. نتایج نشان داد که شاخص‌های نوین نقدینگی در ارزیابی عملکرد مالی شرکت‌ها تصویر دقیق‌تری نسبت به شاخص‌های سنتی نقدینگی به استفاده‌کنندگان از اطلاعات مالی در جهت اخذ تصمیمات بهینه ارائه می‌دهند [۲۰].

فروغی و مندجمی (۲۰۱۱) از درخت تصمیم C&R برای پیش‌بینی ورشکستگی مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران بین سال‌های ۱۹۹۶ تا ۲۰۰۹ استفاده کرده است. متغیرهای پیش‌بینی‌کننده شامل ۲۵ نسبت مالی به‌دست‌آمده از صورت‌های مالی است که در ۴ گروه نسبت‌های نقدینگی، اهرمی، کارایی، سوددهی بخش‌بندی شده است. ۷ نسبت مالی سود عملیاتی به بهره، سود عملیاتی به کل دارایی‌ها، جریان نقد به بدهی‌های بلندمدت، کل بدهی به کل دارایی‌ها، سود خالص به کل دارایی‌ها و دارایی جاری به فروش متغیرهای

تاثیرگذاری بر ورشکستگی شناسایی شدند. دقت آموزش مدل ۹۴/۵ درصد و دقت تست آن ۹۰ درصد گزارش شد [۷].

آذر و کریمی (۱۳۸۹) در تحقیق خود به پیش‌بینی بازده سهام با استفاده از نسبت‌های حسابداری با رویکرد شبکه‌های عصبی پرداختند. در این تحقیق توانایی پیش‌بینی بازده سهام با استفاده از نسبت‌های حسابداری با دو رویکرد شبکه‌های عصبی مصنوعی و رگرسیون حداقل مربعات مورد بررسی قرار گرفته است. متغیرهای مستقل در این تحقیق نسبت‌های حسابداری و متغیر وابسته بازده سهام است؛ بدین منظور نسبت‌های حسابداری برای دو صنعت سیمان و دارو به مدت ۸ سال جمع‌آوری شد. نتایج نشان می‌دهد که با استفاده از نسبت‌های حسابداری تا حدودی می‌توان بازده سهام را پیش‌بینی کرد و هر دو روش رگرسیون حداقل مربعات و شبکه‌های عصبی مصنوعی توانایی پیش‌بینی بازده سهام را دارند؛ اما عملکرد شبکه‌های عصبی مصنوعی از رگرسیون حداقل مربعات بهتر است [۱].

بهرامفر و ساعی (۱۳۸۵) در تحقیق خود ۲۴۱ شرکت پذیرفته‌شده در بورس اوراق بهادار تهران را که در سال‌های بین ۱۳۷۴ تا ۱۳۸۲ سه سال متوالی فعالیت داشته‌اند را مورد بررسی قرار دادند. آن‌ها نشان دادند که نسبت‌های مرتبط با سنجش فعالیت و نسبت‌های مرتبط با اندازه‌گیری وضعیت بدهی‌ها، اندازه شرکت و نوع صنعت در پیش‌بینی رتبه عملکرد شرکت‌ها مفید است. با استفاده از تحلیل آماری لاجیت احتمال تعلق یک شرکت به گروه شرکت‌های موفق و ناموفق را از نظر بازده سهام (عملکرد بازار) و بازده حقوق صاحبان سهام (عملکرد مالی) پیش‌بینی کردند. یافته‌های تحقیق نشان داده صحت پیش‌بینی مدل‌های تعیین‌شده برای پیش‌بینی رتبه عملکرد بازده سهام و عملکرد بازده حقوق صاحبان سهام، چه در نمونه تعیین مدل و چه در نمونه کنترل آن، از نظر آماری با اهمیت هستند و آماره‌های مدل پیش‌بینی بازده حقوق صاحبان سهام، بهتر از مدل پیش‌بینی بازده سهام است [۲].

گودرزی (۱۳۸۴) به تحقیقی با عنوان «بررسی ارتباط بین نسبت‌های مالی با بازده سهام در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران» پرداخت. نتایج تحقیق حاکی از وجود ارتباط بین نسبت‌های جاری، گردش دارایی‌های ثابت، قدرت سودآوری و بازده فروش با بازده سهام است [۸].

مهرانی و همکاران (۱۳۸۳) با استفاده از اطلاعات دو ساله ۱۳۸۱ و ۱۳۸۲ ۱۲۰ شرکت و ۶ متغیر اقدام به بررسی و پیش‌بینی شرکت‌های موفق و ناموفق کرده‌اند. متغیرهای مورد بررسی آن‌ها متشکل از ۳ متغیر مالی و ۳ متغیر غیرمالی است که عبارت‌اند از بازده حقوق صاحبان سهام، رشد فروش، رشد سود، حجم معاملات، تعداد خریداران و تعداد دفعات معامله. آن‌ها شرکت‌های موفق را به‌گونه‌ای تعریف کردند که باید حداقل ۵ نسبت از نسبت‌های یاد شده در

سال ۱۳۸۲ نسبت به سال ۱۳۸۱ افزایش داشته باشد و شرکت‌های ناموفق نیز باید حداقل ۵ نسبت آن‌ها در سال ۱۳۸۲ کاهش داشته باشد. آن‌ها تفاوت معناداری بین بازده سهام این دو گروه پیدا کردند. در مدل نهایی آن‌ها، تنها دو متغیر بازده حقوق صاحبان سهام و رشد فروش به‌عنوان متغیرهای توضیح‌دهنده شرکت‌های موفق و ناموفق استفاده شده است و سایر متغیرها رابطه مهمی با طبقه‌بندی انجام‌شده نداشته‌اند [۲۱].

نمازی و رستمی (۱۳۸۵) به بررسی رابطه بین نسبت‌های مالی و میزان بازده سهام شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران پرداختند. دوره زمانی این تحقیق ۱۳۷۸-۱۳۸۲ است. برای بررسی صحت فرضیه‌های تحقیق از روش رگرسیون ترکیبی (تلفیق داده‌ها زمانی و مقطعی) استفاده شده است. نمونه انتخابی تحقیق شامل ۶۳ شرکت است. نسبت‌های مالی مورد استفاده به‌عنوان متغیر مستقل در پنج گروه قرار گرفتند. از آنجاکه به‌طور معمول تاثیر تغییرات متغیرهای مستقل بر روی متغیر وابسته آنی نیست و به دلیل ناکارای بودن بورس اوراق بهادار تهران، محققان تاثیر اطلاعات مالی هر سال را بر بازده سهام در سال‌های بعد اندازه‌گیری کردند، یعنی متغیرهای مستقل را به‌صورت تاخیری وارد مدل کردند. نتایج این تحقیق نشانگر وجود رابطه معناداری بین کلیه نسبت‌های مالی و نرخ بازده سهام است [۲۳].

۳. روش‌شناسی تحقیق

این تحقیق بر اساس هدف از نوع توسعه‌ای و مدل‌سازی است و در گام بعدی که ارزیابی و اعتبارسنجی در آن انجام می‌شود؛ بر اساس هدف از نوع کاربردی و از لحاظ گردآوری اطلاعات از نوع پیمایشی است. اطلاعات مورد استفاده برای اجرای پژوهش با بهره‌گیری از بانک‌های اطلاعاتی سیستمی، روش‌های کتابخانه‌ای به‌دست آمده است. برای توسعه پروژه‌های داده‌کاوی رویه‌های مختلفی توسعه داده شده است، در این پژوهش روش استاندارد داده‌کاوی کریسپ دی ام استفاده شده است. این مدل فعالیت‌های داده‌کاوی را به شش مرحله (شناخت کسب‌وکار، شناخت داده، آماده‌سازی داده، مدل‌سازی، ارزیابی و توسعه) دسته‌بندی می‌کند. به‌منظور ارزیابی مدل‌ها از ۳ روش معیارهای به‌دست‌آمده از ماتریس اغتشاش، نمودار بهره، اعتبارسنجی متقابل استفاده می‌شود. متغیرهای مورد بررسی در این پژوهش در جدول ۱ آورده شده است.

نسبت بازده دارایی‌ها (ROA) و حقوق صاحبان سهام (ROE) به ۵ گروه با عملکرد خیلی بد، بد، متوسط، خوب و خیلی خوب تقسیم‌بندی شده است و به‌عنوان متغیر نشان‌دهنده عملکرد در نظر گرفته می‌شود. با به‌کارگیری ۴ درخت تصمیم بهترین الگوریتم از لحاظ ارزیابی مشخص می‌شود.

جدول ۱. متغیرهای مستقل و وابسته مورد بررسی در پژوهش

کد متغیر	متغیرهای مستقل	شرح نسبت‌ها
A	نسبت‌های نقدینگی	
A1	نسبت جاری	جمع دارایی‌های جاری به جمع بدهی‌های جاری
A2	نسبت آنی	(جمع دارایی‌های جاری - موجودی مواد و کالا + پیش پرداخت‌ها و سفارشات) به جمع بدهی‌ها جاری
A3	نسبت نقدینگی	(سرمایه‌گذاری‌های کوتاه‌مدت + وجوه نقد و موجودی نزد بانک) به جمع بدهی‌های جاری
A4	نسبت دارایی جاری	جمع دارایی‌های جاری به جمع کل دارایی‌ها
A5	نسبت کفایت نقد	جریان خالص ورود و خروج وجوه نقد به (بازپرداخت بدهی‌های بلندمدت + خرید سایر دارایی‌ها + خرید دارایی ثابت + سود پرداختنی بابت تسهیلات + سود سهام پرداختنی)
A6	نسبت گردش نقد	جریان خالص ورود و خروج وجوه نقد به جمع بدهی‌های جاری
B	نسبت‌های بدهی	
B1	نسبت بدهی	جمع کل بدهی‌ها به جمع کل دارایی‌ها
B2	نسبت بدهی به ارزش ویژه	جمع کل بدهی‌ها به حقوق صاحبان سهام
B4	نسبت مالکانه	حقوق صاحبان سهام به جمع کل دارایی‌ها
B5	نسبت بدهی بلندمدت به ارزش ویژه	بدهی‌های بلندمدت به حقوق صاحبان سهام
B6	نسبت بدهی جاری به ارزش ویژه	جمع بدهی‌های جاری به حقوق صاحبان سهام
B7	نسبت دارایی ثابت به ارزش ویژه	خالص دارایی‌های ثابت به حقوق صاحبان سهام
C	نسبت‌های کارایی	
C1	گردش دارایی ثابت	جمع درآمدها به خالص دارایی‌های ثابت
C2	گردش مجموع دارایی‌ها	جمع درآمدها به جمع کل دارایی‌های جاری
C3	گردش سرمایه جاری	جمع درآمدها به (جمع دارایی‌های جاری - جمع بدهی‌های جاری)
C4	نسبت کالا به سرمایه در گردش	موجودی مواد و کالا به (جمع دارایی‌ها جاری - جمع بدهی‌های جاری)
D	نسبت‌های سودآوری	
D1	سود خالص به فروش	سود خالص به جمع درآمدها
D2	سود ناخالص به فروش	سود (زیان) قبل از کسر مالیات به جمع درآمدها
D3	سود عملیاتی به جمع درآمدها	سود (زیان) عملیاتی به جمع درآمدها
D4	سود ناویژه به فروش	سود (زیان) ناویژه به جمع درآمدها
D5	سود خالص به سود ناویژه	سود خالص به سود (زیان) ناویژه
E	متغیرهای وابسته	شرح نسبت‌ها
E1	بازدهی حقوق صاحبان سهام (ROE)	سود خالص به حقوق صاحبان سهام
E2	بازده دارایی‌ها (ROA)	سود خالص به دارایی‌ها

اطلاعات مجموعه‌ای از نسبت‌های به‌دست‌آمده از صورت‌های مالی مربوط به شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران است و با استفاده از نرم‌افزار رهاورد نوین جمع‌آوری شده و با استفاده از نرم‌افزارهای SPSS Statistic، SPSS Modeler و اکسل مراحل آماده‌سازی و مدل‌سازی و ارزیابی انجام شده است.

قلمرو زمانی تحقیق با توجه به نسبت‌های به‌دست‌آمده از صورت‌های مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران در سال‌های ۱۳۹۳-۱۳۹۰ است.

جامعه آماری در این تحقیق شامل ۵۳۴ شرکت پذیرفته‌شده در بورس اوراق بهادار تهران در سال‌های ۱۳۹۳-۱۳۹۰ است. نمونه‌گیری به‌صورت حذفی انجام می‌شود، بدین ترتیب که کلیه شرکت‌های عضو جامعه آماری تنها در صورت تحقق شرایط زیر در نمونه لحاظ شده و در غیر این صورت از نمونه حذف می‌شوند:

- اطلاعات مالی شرکت طی سال‌های ۱۳۹۳-۱۳۹۰ در دسترس باشد.

- شرکت‌هایی که تا قبل از سال ۱۳۹۰ در بورس اوراق بهادار پذیرفته شده باشند و نام شرکت در دوره مورد بررسی از بین شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران حذف نشده باشد.

۴. تحلیل داده‌ها

فرایند ساخت مدل. شناسایی و آماده‌سازی داده‌ها در این تحقیق، داده‌ها به روش مشاهده‌ای جمع‌آوری شده است. اطلاعات از صورت‌های مالی و با استفاده از نرم‌افزار رهاورد نوین به‌دست آمده است. با توجه به ماهیت نسبت‌های مالی داده‌ها همگی از نوع داده‌های نسبتی و این متغیرهای کمی با عدد قابل سنجش هستند.

معمولاً این گام از فرایند داده‌کاوی و کشف دانش طولانی‌تر و مهم‌تر محسوب می‌شود؛ به‌طوری‌که هر قدر کیفیت آماده‌سازی داده‌ها بهتر انجام شود و زمان بیشتری برای آن صرف شود، مدل‌سازی و نتیجه‌گیری نیز بهتر و آسان‌تر خواهد بود و نتایج اعتبار بیشتری خواهند داشت. متغیرهای بر اساس نتایج آزمون کولموگروف-اسمیرنوف نرمال نبوده است. برای بررسی بین همبستگی بین متغیرها از آزمون پیرسون استفاده شده است.

بر اساس ضرایب همبستگی به‌دست‌آمده، متغیرهای نسبت آنی و نسبت جاری و همچنین نسبت بدهی به ارزش ویژه، نسبت مالکانه و نسبت بدهی و همچنین سود خالص به فروش و سود ناخالص به فروش دارای شدت همبستگی بالایی هستند (ضریب همبستگی پیرسون بزرگتر از ۰/۸) هستند. با توجه به ماهیت استفاده از متغیرها، نسبت‌های آنی و نسبت سود ناخالص به فروش و نسبت بدهی به ارزش ویژه از میان متغیرها حذف شد و ۱۹ متغیر، به‌عنوان متغیرهای مستقل درخت تصمیم باقی ماند.

داده‌های بدون مقدار حذف و داده‌های خارج از محدوده با استفاده از فرمول شوارتز شناسایی و مدیریت شدند؛ همچنین به‌منظور هم مقیاس کردن داده‌های عددی تمام نسبت‌ها در بازه ۱- تا ۱+ نرمال شدند.

مدل‌سازی. تمامی مدل‌ها برای متغیر هدف بازده دارایی‌ها و بازده حقوق صاحبان سهام به‌طور جداگانه اجرا شده است. بر اساس ارزیابی Holdout در ابتدا ۷۰ درصد از داده‌ها برای آموزش و ۳۰ درصد آن‌ها برای تست در نظر گرفته می‌شوند و مدل‌ها بر اساس داده‌های آموزش ساخته می‌شوند و با داده‌های تست مورد آزمون نهایی قرار می‌گیرند.

ارزیابی مدل. دانشی که در مرحله یادگیری مدل تولید می‌شود، باید در مرحله ارزیابی مورد تحلیل قرار گیرد تا بتوان ارزش آن را تعیین نمود و در پی آن کارایی الگوریتم یادگیرنده مدل را نیز مشخص کرد. این معیارها را می‌توان هم برای مجموعه داده‌های آموزشی در مرحله یادگیری و هم برای مجموعه رکوردهای آزمایشی در مرحله ارزیابی محاسبه نمود؛ همچنین لازمه موفقیت در بهره‌مندی از علم داده‌کاوی تفسیر دانش تولید و ارزیابی شده است. بهترین مدل از لحاظ عملکرد در این بخش مورد بررسی قرار می‌گیرد.

ماتریس اغتشاش. بر اساس ماتریس اغتشاش به‌دست‌آمده در هر مدل، دو شاخصه صحت و نرخ خطا به‌دست آمده است. نتایج هر یک از الگوریتم‌ها در جداول ۲ و ۳ آورده شده است.

جدول ۲. صحت به‌دست‌آمده در پیش‌بینی متغیر هدف بازده دارایی‌ها

متغیر هدف بازده دارایی‌ها				
نوع مدل	آموزشی		آزمایشی	
QUEST	درست	۶۳۰	%۶۵/۸۳	۲۳۷
	اشتباه	۳۲۷	%۳۴/۱۷	۱۴۹
C5.0	درست	۹۲۳	%۹۶/۴۵	۳۲۲
	اشتباه	۳۴	%۲/۵۵	۶۴
CHAID	درست	۸۵۱	%۸۸/۹۲	۳۱۹
	اشتباه	۱۰۶	%۱۱/۰۸	۶۷
C&R	درست	۷۲۵	%۷۵/۷۶	۲۶۶
	اشتباه	۲۳۲	%۲۴/۲۴	۱۲۰

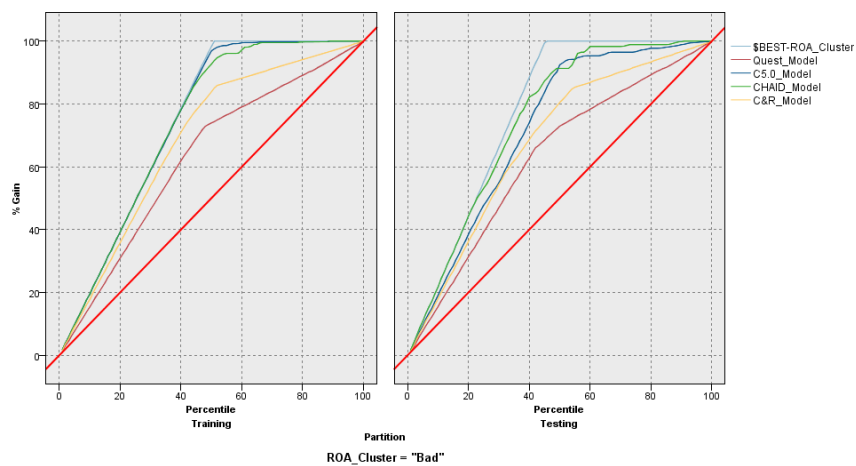
با توجه به مقادیر به‌دست‌آمده به ترتیب درخت تصمیم C5.0، CHAID، C&R و QUEST با صحت کلی ۸۳/۴۲، ۸۲/۶۴، ۶۸/۹۱ و ۶۱/۴۰ دارای بیشترین شاخص ارزیابی صحت هستند.

جدول ۳. صحت به‌دست‌آمده در پیش‌بینی متغیر هدف بازده حقوق صاحبان سهام

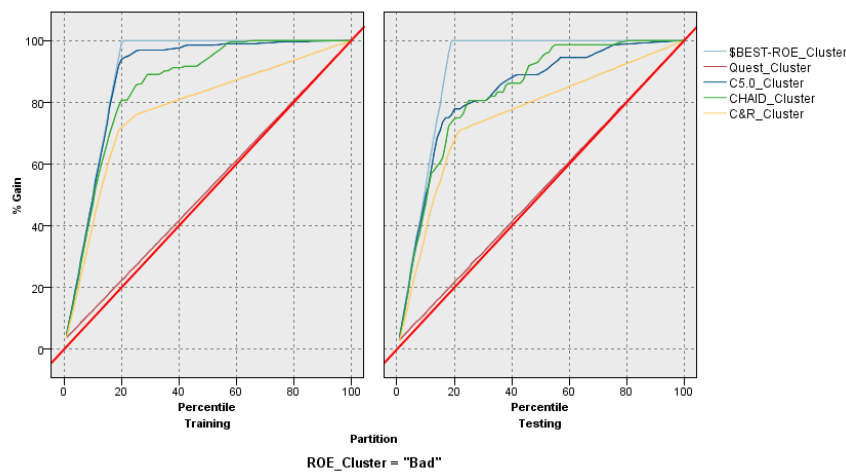
متغیر هدف حقوق صاحبان سهام			
نوع مدل	آموزشی	آزمایشی	
QUEST	درست	۶۹۳	٪۷۲/۴۱
	اشتباه	۲۶۴	٪۲۷/۵۹
C5.0	درست	۹۲۴	٪۹۶/۵۵
	اشتباه	۳۳	٪۳/۴۵
CHAID	درست	۸۴۱	٪۸۷/۸۸
	اشتباه	۱۱۶	٪۱۲/۱۲
C&R	درست	۷۹۲	٪۸۲/۷۶
	اشتباه	۱۶۵	٪۱۷/۲۴

بر اساس صحت و نرخ خطا در مدل‌ها به ترتیب C5.0، C&R، CHAID و QUEST با مقادیر ۸۵/۴۹، ۸۲/۱۲، ۸۲/۱۲، ۷۵/۵۶ است. با توجه به مقادیر به‌دست‌آمده مدل C5.0 از قابلیت اتکای بیشتری برخوردار است؛ البته با توجه به مقادیر بالای ۷۰ درصد می‌توان گفت با تقریبی خوبی تمامی مدل‌ها از قابلیت پیش‌بینی خوبی برخوردار هستند.

نمودار بهره. یکی دیگر از معیار ارزیابی استفاده از نمودار بهره است. بهره به‌صورت نسبتی از کل موفقیت‌ها است که در هر چندک رخ می‌دهد و به‌صورت (تعداد موفقیت‌ها در چندک‌ها/کل تعداد موفقیت‌ها) * ۱۰۰٪ تعریف می‌شود. نتایج نمودار بهره نیز همراستا با ارزیابی شاخصه‌ها است. با توجه به اشکال ۱ و ۲ دو درخت C5.0 و CHAID و در بیشتر صدک‌ها از عملکرد بهتری نسبت به دو درخت تصمیم دیگر برخوردار بوده است.

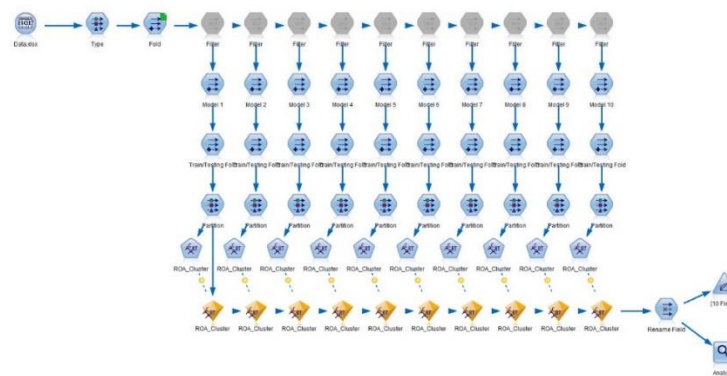


شکل ۱. نسبت صدک در موفقیت در متغیر هدف بازده دارایی‌ها



شکل ۲. نسبت صدک در موفقیت در متغیر هدف حقوق صاحبان سهام

اعتبار سنجی متقابل. این روش نمونه‌ها را به K زیرمجموعه هم اندازه تقسیم می‌کند و یک زیرمجموعه را گروه آزمایش و دیگر زیرمجموعه‌ها را گروه یادگیری در نظر می‌گیرد. این کار تکرار می‌شود تا هر زیرمجموعه یک بار به‌عنوان گروه آزمایش و یک بار به‌عنوان گروه یادگیری انتخاب شود. در این پژوهش از ۱۰ تکرار برای اعتبارسنجی استفاده شده است. شکل ۳ پیاده‌سازی روش ۱۰-fold را برای الگوریتم C&R نشان می‌دهد. نتایج این گروه‌بندی به‌عنوان معیاری برای بررسی کیفیت قوانین حاصل و اعتبارسنجی مدل به‌کار می‌رود [۱۷، ۲۸]. جداول ۴ و ۵ خلاصه‌شده ده مرحله انجام‌شده را نشان می‌دهند.



شکل ۳. نمایی از اجرای ارزیابی ۱۰-fold برای درخت تصمیم C&R

جدول ۴ نتایج اعتبارسنجی متقابل را برای متغیر هدف بازده دارایی‌ها نشان می‌دهد. در بین داده‌های آموزشی، بیشترین دقت مربوط به مدل C5.0 است و بعد از آن CHAID، C&R و QUEST می‌باشد که نتایج همراستا با ارزیابی مدل بر اساس شاخص صحت و نمودار بهره است. در داده‌های آزمایشی هم الگوریتم C5.0 دارای متوسط دقت بالای ۷۰٪ است و نشان‌دهنده این موضوع است که دقت قوانین به‌دست‌آمده در این مدل از سایر مدل‌ها بیشتر است.

جدول ۴. نتایج اعتبارسنجی متقابل در متغیر هدف بازده دارایی‌ها

نوع مدل	داده‌های آموزشی			داده‌های آزمایشی		
	متوسط دقت	حداکثر دقت	حداقل دقت	متوسط دقت	حداکثر دقت	حداقل دقت
CHAID	٪۶۷/۸۷	٪۷۰/۴۵	٪۶۴/۵۷	٪۶۵/۱۳	٪۷۱/۰۵	٪۵۹/۸۷
C5.0	٪۷۳/۵۴	٪۸۳/۹۱	٪۶۶/۸۹	٪۷۸/۳۷	٪۸۴/۶۷	٪۶۸/۶۷
QUEST	٪۵۷/۷۸	٪۶۶/۱۷	٪۴۹/۴۳	٪۶۴/۰۰	٪۷۵/۲۰	٪۵۶/۰۰
C&R	٪۶۵/۶۸	٪۷۰/۲۶	٪۴۹/۰۹	٪۶۸/۹۹	٪۷۵/۱۹	٪۴۹/۶۱

نتایج اعتبارسنجی متقابل برای متغیر هدف بازده حقوق صاحبان سهام نیز در جدول ۵ مشخص شده است، در بین داده‌های آزمایشی، بیشترین دقت همانند حالتی که متغیر هدف بازده دارایی‌ها است، مدل C5.0 و بعد از آن CHAID، C&R و QUEST است. در داده‌های آزمایشی متوسط دقت تمامی مدل‌ها حدوداً از ۶۰ درصد بالاتر است.

جدول ۵

نوع مدل	داده‌های آموزشی			داده‌های آزمایشی		
	متوسط دقت	حداکثر دقت	حداقل دقت	متوسط دقت	حداکثر دقت	حداقل دقت
CHAID	٪۶۷/۶۸	٪۷۰/۵۴	٪۶۴/۴۰	٪۶۶/۵۲	٪۷۱/۷۴	٪۵۴/۳۵
C5.0	٪۷۳/۹۰	٪۸۳/۵۷	٪۶۶/۷۲	٪۷۵/۶۱	٪۸۷/۸۸	٪۷۰/۴۵
QUEST	٪۵۸/۲۸	٪۶۶/۲۸	٪۵۰/۵۴	٪۵۹/۱۵	٪۷۰/۱۰۰	٪۴۷/۶۹
C&R	٪۶۵/۷۷	٪۷۰/۵۲	٪۴۸/۵۷	٪۶۸/۲۸	٪۷۳/۷۷	٪۵۴/۹۲

۵. بحث و نتیجه‌گیری

در این تحقیق با توجه به این واقعیت که پیش‌بینی عملکرد از دیدگاه بازده دارایی‌ها و بازده حقوق صاحبان سهام مورد علاقه استفاده‌کنندگان و افراد ذی‌نفع از گزارشات مالی است. دو نسبت بازده دارایی‌ها و حقوق صاحبان سهام به‌عنوان دو نمونه از معیارهای ارزیابی عملکرد به‌عنوان متغیرهای وابسته در نظر گرفته شدند. چهار الگوریتم درخت تصمیم شامل درخت تصمیم CHAID، C5.0، QUEST، و C&R برای هر دو متغیر هدف به‌طور مجزا اجرا شد و تمامی آن‌ها با چهار روش ارزیابی مورد سنجش قرار گرفتند. با توجه به یافته‌های پژوهش و تجارب به‌دست‌آمده در مقایسه چهار درخت تصمیم در روش‌های ارزیابی Holdout، اعتبارسنجی متقابل، شاخصه‌های ارزیابی و نمودار بهره، درخت تصمیم C5.0 دارای بیشترین صحت است و نتایج حاکی از آن است که این مدل دارای قدرت پیش‌بینی‌کنندگی مناسبی در زمینه رتبه‌بندی شرکت‌ها از لحاظ عملکرد مالی است. به‌طور کلی مدل‌های به‌دست‌آمده با متغیر هدف حقوق صاحبان سهام نسبت به مدل‌های به‌دست‌آمده در حالتی که متغیر هدف بازده دارایی‌ها است از اعتبار بالاتری برخوردار هستند و در بین مدل‌ها نیز به ترتیب مدل‌های C5.0، C&R، CHAID و QUEST در هر دو حالت بهترین دقت را در ارزیابی‌ها دارا بودند و نشان‌دهنده این موضوع است که دقت قوانین به‌دست‌آمده در این مدل‌ها تا چه اندازه قابل اعتماد است.

منابع

1. Azar, A., & Karimi, S. (2010). Neural Network Forecasts of Stock Return Using Accounting Ratios. *Financial Research*, 11(28), 3-20.
2. Bahramifar, N., & M. J. Saei. (2006). The Prediction of financial and stock market performance using the financial statements data: the case of Tehran stock exchange. *Journal of the accounting and auditing review*, 13(1), 45-70.
3. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. New York: Chapman & Hall/CRC.
4. De Ville, B., & Neville, P. (2013). *Decision Trees for Analytics Using SAS Enterprise Miner*. SAS Institute.
5. Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A Decision tree approach. *Expert System with Application*, 40(10), 3970-3983.
6. Doolatabadi, H. R., Hoseini, S. M., & Tahmasebi, R. (2013). Using decision tree model and logistic regression to predict companies financial bankruptcy in Tehran stock exchanges. *International Journal of Emerging Research in Management & Technology*, 2(9), 7-16.
7. Foroghi, D., & Monadjemi, A. (2011). Applying decision tree to predict bankruptcy. In Computer Science and Automation Engineering (CSAE), *International Conference on*, 4, 165-169. IEEE.
8. Godarzi, A. (2005). Analytical review of the relationship between financial ratios and stock returns of companies listed on the Tehran Stock Exchange, (master dissertation), Tehran, Islamic Azad University, Science and Research Branch.
9. Gorunescu, F. (2011). Data Mining: Concepts, models and techniques, 12. *Springer Science & Business Media*.
10. Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT press.
11. Hashemi, S.A., Hosseini, S.M., & Barandan, S. (2012). The comparison of incremental information content of cash and accrual ratios for financial performance evaluation of companies using data mining. *Financial accounting research*, 2(12), 63-81.
12. Hejazi, R., Mohamadi, SH., Aslani, Z., & Aghajani, M. (2012). Earnings management prediction using neural networks and decision tree in TSE. *Journal of the accounting and auditing review*, 19(68), 31-46.
13. Hosseini, M., & Rashidi, Z. (2013). Using decision tree model and logistic regression to predict companies financial bankruptcy in Tehran Stock Exchanges. *Financial accounting research*, 3(17), 105-130.
14. Jooste, L. (2006). Cash flow ratios as a yardstick for evaluating financial performance in African businesses. *Managerial Finance*, 32(7), 569-576.
15. Kadoya, S., Kuroko, T., & Namatame, T. (2008). Contrarian Investment Strategy with Data envelopment analysis Concept. *European Journal of Operational Research*, 189(1), 120-131.
16. Karimi, T. & Sadeghi Moghadam, M. (2014). *Rough Set and Grey Set Theory: principles, applications and software's*. First edition. Tehran: Mehraban nashr.
17. Karami, G. R., & Talaeei, L. (2013). Predictability of stock returns using financial ratios in the companies listed in Tehran Stock Exchange. *International Research Journal of Applied and Basic Sciences*, 4(12), 4261-4273.

18. Koyuncugil, A. S., & Ozgulbas, N. (2012). Financial early warning system model and data mining application for risk detection. *Expert Systems with Applications*, 39(6), 6238-6253.
19. Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2), 209-235.
20. Mahdavi, G., & Ghorbani, A. (2012). A comparative study of modern and traditional liquidity indicators in assessing the financial performance of companies listed in Tehran Stock Exchange. *Financial accounting research*, 4(11), 67-88.
21. Mehrani, S., Mehrani, K., & Karami, GH. (2004). Using historical information of financial and non-financial firms to differentiate successful and unsuccessful. *Journal of the accounting and auditing review*, 11(38), 77-92.
22. Miner, G., Nisbet, R. & Elder, J., (2009). *Handbook of Statistical Analysis and Data Mining Applications*. 1 ed. s.l.: Academic Press.
23. Namazi, M., & Rostami, N. (2006). Examines the relationship between financial ratios and stock returns of companies listed on the Stock Exchange. *Journal of the accounting and auditing review*, 13(44), 44-105.
24. Olson, D., & Mossman, C. (2003). Neural network forecasts of Canadian stock returns using accounting ratios. *International Journal of Forecasting*, 19(3), 453-465.
25. Salehi, M., & Farokhi pilerudi, L. (2014). Using Neural Network Model and Decision Tree to Predict Earning Managment in Tehran Stock, *Agricultural Economics*. 8, 89-105.
26. Yu, G., & Wenjuan, G. (2010). Decision tree method in financial analysis of listed logistics companies. *International conference on intelligent computation technology and automation*.
27. Wang, H., Jiang, Y., & Wang, H. (2009). Stock return prediction based on Bagging-decision tree. In *2009 IEEE International Conference on Grey Systems and Intelligent Services*: 1575-1580. IEEE.
28. Zhong, N., Dong, J., & Ohsuga, S., (2001). Using Rough Sets with Heuristics for Feature Selection. *Journal of Intelligent Information Systems*, 16(3): 199-214.