

## تعیین مدل ارزیابی عملکرد مالی شرکت‌های بورس اوراق بهادار تهران بر مبنای داده‌کاوی

ایمان رئیسی وانانی\*  
قاسم بولو\*\*  
شهره زرکش (نویسنده مسئول)\*\*\*  
تاریخ ارسال: ۹۵/۸/۱۲  
تاریخ پذیرش: ۹۶/۲/۱۵

### چکیده

تعیین عملکرد شرکت‌ها با استفاده از نسبت‌های مالی همواره جالب توجه بسیاری از محققان و پژوهشگران بوده است. شناسایی عوامل مالی که بیشترین تاثیر را بر عملکرد شرکت‌ها داشته یکی از مباحث مهم برای تصمیم‌گیرندگان است. در این پژوهش با به کارگیری استاندارد داده‌کاوی، ۴ مدل درخت تصمیم کاشف تعاملات خودکار کای دو<sup>۱</sup>، آماری موثر عاری از تعصب و سریع<sup>۲</sup>، خوشه بندی و رگرسیون<sup>۳</sup> و سی فایو<sup>۴</sup> به منظور ارزیابی عملکرد پیاده‌سازی و مدل‌ها با معیارهای ارزیابی مقایسه شدند. همچنین نسبت‌های تاثیرگذار در ارزیابی مالی شرکت‌ها شناسایی شد. ۶۷۳ شرکت بورس اوراق بهادار تهران بر اساس ۳۰ نسبت بدست آمده از صورت‌های مالی سال‌های ۱۳۹۳-۱۳۹۰ بررسی شده است. نتایج حاصل نشان می‌دهد در مقایسه ۴ درخت تصمیم در روش‌های ارزیابی تخمین نمونه آزمایشی، اعتبار سنجی متقابل و شاخص‌های ارزیابی درخت تصمیم کاشف تعاملات خودکار کای دو دارای بیشترین صحت و دقت است. البته مدل‌های دیگر نیز از اطمینان بالایی (بالای ۸۰ درصد) برخوردار هستند و می‌توان از تمامی آن‌ها استفاده کرد. همچنین در نتایج داده‌کاوی حاضر نسبت سود خالص به فروش، نسبت بار مالی وام و گردش مجموع دارایی‌ها در میان نسبت‌های مالی مورد بررسی اهمیت بالاتری در پیش‌بینی عملکرد داشتند.

**مفاهیم کلیدی:** عملکرد مالی، ارزیابی مدل، داده‌کاوی، درخت تصمیم.

---

\* استادیار مدیریت صنعتی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی [imanraeesi@atu.ac.ir](mailto:imanraeesi@atu.ac.ir)  
\*\* دانشیار مدیریت و حسابداری، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی [blue@atu.ac.ir](mailto:blue@atu.ac.ir)  
\*\*\* دانشجوی کارشناسی ارشد، مدیریت مالی، دانشگاه علامه طباطبائی [shohreh.zarkesh@gmail.com](mailto:shohreh.zarkesh@gmail.com)

## مقدمه و بیان مسأله

یکی از مباحث مهم در حوزه مدیریت مالی بررسی قابلیت پیش‌بینی و عوامل موثر بر عملکرد شرکت‌ها است. همواره شرکت‌های بورسی از یک طرف و ذینفعان مالی از طرف دیگر به دنبال ارتقا سطح عملکرد شرکت‌ها هستند به طوری که بتوانند عملکرد آن‌ها را مورد بررسی قرار داده و به پیش‌بینی صحیح و راه‌دستیابی به آن برسند.

با پیشرفت تکنولوژی و توسعه ابزارهای قدرتمند برای جمع‌آوری و نگهداری داده‌ها، ابزارهای مدیریت سنتی برای آنالیز داده‌ها کافی نیستند و بدین منظور برای تبدیل این داده‌ها به دانش سازمان یافته ابزارهای پیش از پیش احساس می‌شود. لذا پژوهش حاضر با به کارگیری از مباحث داده‌کاوی به بررسی و ارزیابی نسبت‌های مالی موثر بر عملکرد شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌پردازد. داده‌کاوی به عنوان یکی از روش‌های قدرتمند در زمینه تحلیل داده‌های خام است که گه‌گاه تحت عنوان فرایند کشف دانش نیز بیان می‌شود. داده‌کاوی در حقیقت کشف ساختارهای جالب توجه، غیر منتظره و با ارزش از داخل مجموعه‌ی وسیعی از داده‌ها می‌باشد (Hand, Mannila & Smyth, 2001). از آنجایی که صورت‌های مالی همواره ابزار اصلی انتقال اطلاعات یک شرکت است، با تمرکز بر روی نسبت‌های مالی، می‌توان عملکرد را بررسی کرد. با تجزیه و تحلیل نتایج بدست آمده امکان بررسی وضعیت کنونی و مقایسه آن‌ها و همچنین درک آینده شرکت از لحاظ عملکرد میسر می‌شود. تشخیص فاکتورهایی که می‌تواند به درستی عملکرد واحد تجاری را پیش‌بینی کند برای هر تصمیم‌گیرنده بسیار جالب است (Delen, Kuzey & Uyar, 2013).

یکی از برتری‌های روش درخت تصمیم انعطاف‌پذیری و قابل فهم بودن آن است (Gorunescu, 2011) به طوری که قابلیت مقایسه و مسیر رسیدن به هدف را به خوبی مشخص می‌کند. بدین منظور در این تحقیق تعیین عملکرد شرکت‌های بورس اوراق بهادار تهران از منظر چهار روش درخت تصمیم مورد ارزیابی قرار می‌گیرد تا در میان روش‌های درخت تصمیم بهترین روش از دیدگاه صحت و دقت مشخص شود و در هر روش نسبت‌هایی که بیشترین تاثیر را در تعیین عملکرد داشته‌اند شناسایی شوند.

### مرور پیشینه تحقیق

استفاده از نسبت‌های مالی برای ارزیابی عملکرد موضوعی جدید محسوب نمی‌شود. در تحقیقات گذشته بیشتر پژوهش‌ها بر اساس تکنیک‌های آماری مانند تحلیل عاملی، رگرسیون خطی و تجزیه و تحلیلی واریانس بوده است. تحقیقات اخیر همسو با پیشرفت بانک‌های اطلاعاتی و ابزارهای فناوری اطلاعات به سمت مباحث داده‌کاوی رفته است. هریگان (۱۹۶۵) ادعا کرد توسعه نسبت‌های مالی محصول منحصر به فرد سیر تکاملی روش‌های حسابداری در آمریکا است و بدین منظور منشا نسبت‌های مالی و به کارگیری آن‌ها به اواخر قرن ۱۹ مربوط می‌شود. نسبت‌های مالی که چکیده‌ی صورت‌های مالی شرکت‌ها محسوب می‌شوند مزایای بسیاری را در بر دارند که می‌توان به ارزیابی عملکرد مدیران به منظور پاداش و یا برنامه ریزی برای آینده و سرمایه گذاری‌های آتی با استفاده از اطلاعات تاریخی موجود اشاره کرد (Ross, Westerfield & Jordan, 2008). از دیگر مزایای نسبت‌های مالی، بکارگیری آن‌ها در مباحث پیش‌بینی عملکرد آتی است. آلتمن (۱۹۶۸) در تحقیقات تجربی خود از نسبت‌های مالی به عنوان متغیرهای ورودی مدل پیش‌بینی کننده بحران مالی استفاده کرد. در حقیقت بیشتر تحقیقات بر روی آنالیز و پیش‌بینی ورشکستگی شرکت‌ها متمرکز شده‌اند تا با ارزیابی عملکرد به این هدف دست پیدا کنند (Kumar & Ravi, 2007). فروغی و مندجمی (۲۰۱۱) از درخت تصمیم C&R برای پیش‌بینی ورشکستگی مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران بین سال‌های ۱۹۹۶ تا ۲۰۰۹ استفاده کرده است. متغیرهای پیش‌بینی کننده شامل ۲۵ نسبت مالی بدست آمده از صورت‌های مالی است که در ۴ گروه نسبت‌های نقدینگی، اهرمی، کارایی، سوددهی بخش بندی شده است. ۷ نسبت مالی سود عملیاتی به بهره، سود عملیاتی به کل دارایی‌ها، جریان نقد به بدهی‌های بلند مدت، 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. در پژوهشی با عنوان "ارزیابی عملکرد واحد تجاری با استفاده از نسبت‌های مالی با رویکرد درخت تصمیم" مقایسه‌ی الگوریتم‌های درخت تصمیم انجام و با استفاده از نسبت‌های مالی، عملکرد شرکت‌های بورس استانبول را بررسی شده است (Delen, et al, 2013). نتایج نشان می‌دهد نسبت سود قبل از مالیات به حقوق صاحبان سهام و حاشیه سود خالص دو متغیر اساسی در

پیش‌بینی عملکرد شرکت‌ها هستند.

کرمی و طلایی (۲۰۱۳) ارتباط بین نسبت‌های قیمت به سود، ارزش دفتری به ارزش بازار، نرخ سود نقدی و بازده سرمایه‌ی 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. در سال ۲۰۱۳ دولت آبادی و همکاران با استفاده از مدل درخت تصمیم و رگرسیون لاجستیک به پیش‌بینی ورشکستگی مالی شرکت‌های بورسی در بازه‌ی 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. یو و ونجوان (۲۰۱۰) برای بررسی اینکه کدام یک از نسبت‌های مالی قوی‌ترین تاثیر را بر رشد سود شرکت‌های تدارکاتی را دارند، از روش C5.0 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. 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. 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 لاجستیک پرداختند. نمونه آماری شمال ۱۹۹ شرکت از بین شرکت‌های پذیرفته شده 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. تحقیقات بررسی شده به طور خلاصه در جدول ۱ آورده شده است. با توجه به موضوعات بررسی شده، به طور کلی نسبت‌های مالی همواره راه مناسبی برای بررسی روند مالی یک شرکت بوده است اما بیشتر تحقیقات در حوزه درخت تصمیم حول موضوعات ورشکستگی است و توجه کمتری به مقایسه این روش‌ها و یافتن نسبت‌های مالی تاثیرگذار در تعیین عملکرد مالی شرکت‌ها شده است و این موضوع می‌تواند درخور توجه باشد. ب

جدول ۱. خلاصه‌ی تحقیقات انجام شده بر اساس موضوعات تحقیق

شماره	عنوان تحقیق	نویسندگان تحقیق	موضوع تحقیق
۱	Some empirical bases of financial ratio analysis	Horrigan, 1965	توسعه نسبت‌های مالی محصول منحصر به فرد سیر تکاملی روش‌های حسابداری در آمریکا است
۲	Fundamentals of corporate finance	Ross, Westerfield & Jordan, 2008	استفاده از نسبت‌های مالی برای ارزیابی عملکرد
۳	Financial ratios, discriminant analysis and the prediction of corporate bankruptcy	Altman, 1968	استفاده از نسبت‌های مالی به عنوان متغیرهای ورودی مدل پیش‌بینی کننده بحران مالی
۴	Bankruptcy prediction in banks and firms via statistical and intelligent techniques-A review	Kumar & Ravi, 2007	تمرکز بر آنالیز و پیش‌بینی ورشکستگی شرکت‌ها به منظور دستیابی به ارزیابی عملکرد
۵	Applying decision tree to predict bankruptcy	Foroghi & Monadjemi, 2011	شناسایی نسبت‌های مالی تاثیر گذار در پیش‌بینی ورشکستگی
۶	Measuring firm performance using financial ratios	Delen, Kuzey, & Uyar, 2013	مقایسه‌ی الگوریتم‌های درخت تصمیم انجام و بررسی عملکرد شرکت‌های بورس استانبول بر اساس نسبت‌های مالی
۷	Predictability of stock returns using financial ratios in the companies listed in Tehran Stock Exchange	Karami, & Talaei, 2013	استفاده از نسبت‌های مالی به منظور بررسی توانایی آن‌ها در پیش‌بینی بازده سهام
۸	Using decision tree model and logistic regression to predict companies financial bankruptcy in Tehran stock exchanges	Doolatabadi, Hoseini, & Tahmasebi, 2013	استفاده از مدل درخت تصمیم و رگرسیون لجستیک برای پیش‌بینی ورشکستگی مالی شرکت‌های بورس
۹	Decision tree method in financial analysis of listed logistics companies	Yu, & Wenjuan, 2010	استفاده از درخت تصمیم C5.0 به منظور یافتن تاثیر گذارترین نسبت‌های مالی قوی‌ترین تاثیر را بر رشد سود شرکت‌های تدارکاتی
۱۰	Stock return prediction based on Bagging-decision tree	Wang, & Jiang, 2009	بررسی درخت تصمیم و نسبت مالی برای پیش‌بینی سود سال آتی
۱۱	مقایسه محتوای فزاینده اطلاعاتی نسبت‌های نقدی و تعهدی برای عملکرد مالی شرکت‌ها با رویکردی داده‌کاوی	هاشمی، حسینی، و برعدان ۱۳۹۱	مقایسه محتوای فزاینده اطلاعاتی نسبت‌های نقدی و تعهدی برای ارزیابی عملکرد مالی واحدهای تجاری با استفاده از مدل درخت تصمیم کارت و رگرسیون لجستیک

## روش پژوهش

یکی از رویکردهای آنالیز و تحلیل داده‌ها، رویکرد داده‌کاوی است که به کشف دانش می‌پردازد. برای توسعه پروژه‌های داده‌کاوی رویه‌های مختلفی توسعه داده شده است، در این پژوهش روش استاندارد داده‌کاوی کریسپ دی ام<sup>۵</sup> استفاده شده است. این مدل فعالیت‌های داده‌کاوی را به شش مرحله (فهم تجاری، فهم داده، آماده‌سازی داده، مدل‌سازی، ارزیابی و پیاده‌سازی) دسته‌بندی می‌کند.

در مرحله نخست با فهم موضوع به آماده‌سازی و شناخت داده‌ها پرداخته می‌شود. سپس ۴ مدل درخت تصمیم به منظور ارزیابی عملکرد مالی شرکت‌ها ساخته می‌شود؛ بر اساس هدف از نوع توسعه‌ای و مدل‌سازی است و در گام بعدی که با استفاده از 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 بدست آمده است. به منظور ارزیابی مدل‌ها از ۳ روش معیارهای بدست آمده از ماتریس اغتشاش، اعتبارسنجی متقابل و تخمین نمونه آزمایشی استفاده می‌شود.

یکی از اهداف مهم در این پژوهش استفاده از طیف وسیعی از داده‌ها و نسبت‌های مالی بورس اوراق بهادار تهران است؛ تا بتوان با استفاده از داده‌کاوی بهترین مدل بدست آید. از ۲۹ متغیر مستقل، نسبت‌های بازده دارایی ثابت، بازده سرمایه در گردش، سرمایه در گردش خالص، گردش موجودی کالا و دوره وصول مطالبات با توجه به نبود اطلاعات کافی به دست آمده از مجموع متغیرها حذف شدند. همچنین متغیرهای درصد بازده سرمایه و سنجش سودمندی وام با توجه به همبستگی بالا با متغیر وابسته از گروه متغیرهای مستقل خارج شدند. بدین ترتیب ۲۳ متغیر برای ورود به مدل‌ها باقی ماند.

نسبت بازده دارایی‌ها (ROA) به ۲ گروه با عملکرد موفق و ناموفق تقسیم بندی شده است و به عنوان متغیر نشان دهنده‌ی عملکرد در نظر گرفته می‌شود. با بکارگیری ۴ درخت تصمیم تاثیر گذارترین نسبت‌های مالی و بهترین الگوریتم از لحاظ ارزیابی مشخص می‌شود.

جامعه آماری در این تحقیق شامل ۶۷۳ شرکت پذیرفته شده در بورس اوراق بهادار تهران در سال‌های ۱۳۹۳-۱۳۹۰ است. نمونه‌گیری به صورت 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 می‌گردند:

- اطلاعات مالی شرکت طی سال‌های ۱۳۹۳-۱۳۹۰ در دسترس باشد.
- شرکت‌هایی که تا پایان سال ۱۳۹۳ در بورس اوراق بهادار پذیرفته شده باشند و نام شرکت در دوره مورد بررسی از بین شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران حذف نشده باشد.

## فرایند ساخت مدل

### شناسایی و آماده سازی داده‌ها

در این تحقیق، داده‌ها به روش مشاهده‌ای جمع آوری شده است. اطلاعات از صورت‌های مالی و با استفاده از نرم‌افزار رهاورد نوین بدست آمده است. با توجه به ماهیت نسبت‌های مالی داده‌ها همگی از نوع داده‌های نسبتی می‌باشند و این متغیرهای کمی با عدد قابل سنجش هستند.

معمولاً این گام از فرایند داده‌کاوی و کشف دانش طولانی‌تر و مهم‌تر محسوب می‌شود. به طوری که 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 حاصل اعتبار بیشتری خواهند داشت. در این بخش متغیرهای مورد بررسی جمع‌آوری، محاسبه و داده‌های با همبستگی بالا حذف شدند. در این بخش دو نسبت آنی و نسبت سود خالص به فروش به ترتیب با نسبت‌های جاری و سود ناخالص به فروش دارای ضریب همبستگی نزدیک به ۱ هستند و از متغیرهای مورد بررسی حذف شدند.

با توجه به مدل ساخته شده در نرم افزار اس پی اس<sup>۶</sup> داده‌های بدون مقدار حذف و داده‌های خارج از محدوده با استفاده از فرمول شوارتز شناسایی و مدیریت شدند. در گام بعدی به منظور هم مقیاس کردن داده‌های عددی تمام نسبت‌ها در بازه ۱- تا ۱+ نرمال شدند.

### مدل سازی

درخت تصمیم روشی برای نمایش مجموعه‌ای از قوانین است که منتهی به یک رده یا

مقدار می‌شود. مولفه‌های اساسی یک درخت تصمیم عبارتند از: گره تصمیم، شاخه و برگ. بر اساس نوع الگوریتم انتخابی به منظور اجرای درخت تصمیم، درخت ممکن است دو یا تعداد بیشتری شاخه داشته باشد. درخت‌های تصمیم از طریق جداسازی متوالی داده‌ها به گروه‌های مجزا تقسیم می‌شوند که هدف در این فرایند افزایش فاصله بین گروه‌ها در هر جداسازی می‌باشد. یکی از تفاوت‌های موجود بین انواع روش‌های ساخت درخت تصمیم چگونگی اندازه‌گیری فاصله بین گروه‌ها می‌باشد. درختان تصمیم به دو دسته کلی تقسیم می‌شوند، درخت‌های تصمیمی که برای پیش‌بینی متغیرهای گسسته (کیفی) استفاده می‌شوند که درخت‌های طبقه‌بندی نامیده می‌شوند و درخت‌هایی که برای پیش‌بینی متغیرهای پیوسته، استفاده می‌شوند که درخت‌های رگرسیون نام دارند.

با توجه به جدول ۱ و ماهیت نسبت‌های مالی به کار گرفته شده امکان مدل‌سازی از طریق هر ۴ الگوریتم وجود دارد.

کاشف تعاملات خودکار کای دو یک الگوریتم یادگیری با ناظر است ( Miner, Nisbet & Elder, 2009, p.246). این روش یک روش آماری بسیار موثر برای تقسیم بندی، 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 (Miner, Nisbet & Elder, 2009, p.246). همچنین دارای یک مکانیسم توقف است که تصمیم می‌گیرد یک زیر درخت چه هنگام کامل شده است (Miner, Nisbet & Elder, 2009, p.792).

الگوریتم سی فایو نیز یک یادگیرنده با ناظر است که توسط راس کوپنلن 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 زیر مجموعه‌ای را نتوان جداسازی کرد.



جدول ۹. مقایسه الگوریتم‌های درخت تصمیم از جهت ورودی و خروجی (IBM, 2015)

نام الگوریتم	نوع متغیر پیش‌بینی کننده	تعداد شاخه‌های قابل تولید در هر گره	نوع متغیر هدف
C&R	پیوسته، رده بندی، دودویی، ترتیبی و اسمی	دودویی	پیوسته، رده بندی، دودویی، ترتیبی و اسمی
C5,0	پیوسته، رده بندی، دودویی، ترتیبی و اسمی	چند شاخه	دودویی، اسمی و ترتیبی
QUEST	پیوسته، رده بندی، دودویی، ترتیبی و اسمی	دودویی	رده بندی، دودویی و اسمی
CHAID	پیوسته، رده بندی، دودویی، ترتیبی و اسمی	چند شاخه	پیوسته، رده بندی، دودویی، ترتیبی و اسمی

الگوریتم آماری موثر عاری از تعصب و سریع در سال ۱۹۹۷ توسط لو و شی برای متغیرهای پاسخ اسمی طراحی شد. درخت رده بندی 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 این که از مقدار  $p$  برای تصمیم‌گیری استفاده می‌نماید، موجب تشکیل درختی ناریب از متغیرها می‌گردد. این الگوریتم ضمن حفظ دقت برآورد در مدل کارت، از سرعت بالاتری در معرفی یک درخت رده بندی نسبت به آن برخوردار است.

استفاده از درخت خوشه‌بندی و رگرسیون در سال آغاز شد (Beriman, et al, 1984). در مقاله‌ی بریمن و همکاران یک الگوریتم درخت تصمیم دودویی مستعد فرآیند متوالی یا پیش‌بینی کننده مطلق با متغیرهای هدف که به طور بازگشتی کار می‌کند. داده‌ها به دو زیرمجموعه گروه‌بندی می‌شود تا سوابق را همگن‌تر از زیرمجموعه قبلی بکند؛ سپس دو زیرمجموعه دوباره تجزیه می‌شود تا زمانی که معیار همگن بودن یا دیگر معیارهای توقف رضایت بخش باشد. بخش پیش‌بینی کننده 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.

## ارزیابی مدل

بر اساس ارزیابی تخمین نمونه آزمایشی در ابتدا ۷۰ درصد از داده‌ها برای آموزش و ۳۰ درصد آن‌ها برای تست در نظر گرفته می‌شوند. مدل‌ها بر اساس داده‌های آموزش ساخته می‌شوند و با داده‌های تست مورد آزمون نهایی قرار می‌گیرند. داده‌های مورد بررسی پس از مرحله آماده‌سازی به ۲۵۲۸ رکورد تقلیل می‌یابد و با توجه به مدل ارزیابی ۱۷۶۹ رکورد برای آموزش و ۷۵۹ داده برای آزمون مورد استفاده قرار می‌گیرد. با استفاده از مقادیر ماتریس اغتشاش‌های نشان داده شده در جدول ۳ بدست آمده از هر یک از مدل‌ها شاخصه‌های ارزیابی برای داده‌های آموزشی به صورت جدول ۴ خلاصه شده است.

جدول ۳. ماتریس اغتشاش برای هر مدل بر اساس داده‌های آموزشی و آزمایشی

آزمایشی						آموزشی					
صحت کلی			کل	موفق	ناموفق	صحت کلی			کل	موفق	ناموفق
۸۲.۶۱٪	۶۲۷	درست	۳۷۱	۶۴	۳۰۷	۸۰.۵۰٪	۱,۴۲۴	درست	۹۰۶	۱۵۷	۷۴۹
۱۷.۳۹٪	۱۳۲	نادرست	۳۸۸	۳۲۰	۶۸	۱۹.۵۰٪	۳۴۵	نادرست	۸۶۳	۶۷۵	۱۸۸
	۷۵۹	جمع	۷۵۹	۳۸۴	۳۷۵		۱,۷۶۹	جمع	۱۷۶۹	۸۳۲	۹۳۷
				موفق	ناموفق					موفق	ناموفق
۹۰.۳۸٪	۶۸۶	درست	۳۷۱	۳۸	۳۳۳	۹۵.۲۶٪	۱,۶۸۷	درست	۹۰۶	۵۰	۸۵۶
۹.۶۲٪	۷۳	نادرست	۳۸۸	۳۵۲	۳۵	۴.۶۴٪	۸۲	نادرست	۸۶۳	۸۳۹	۳۲
	۷۵۹	جمع	۷۵۹	۳۹۱	۳۶۸		۱,۷۶۹	جمع	۱۷۶۹	۸۸۱	۸۸۸
				موفق	ناموفق					موفق	ناموفق
۹۱.۸۳٪	۶۹۷	درست	۳۷۱	۴۲	۳۲۹	۹۳.۷۸٪	۱,۶۵۹	درست	۹۰۶	۷۵	۸۳۱
۸.۱۷٪	۶۲	نادرست	۳۸۸	۳۶۸	۲۰	۶.۲۲٪	۱۱۰	نادرست	۸۶۳	۸۲۸	۳۵
	۷۵۹	جمع	۷۵۹	۴۱۰	۳۴۹		۱,۷۶۹	جمع	۱۷۶۹	۹۰۲	۸۶۶
				موفق	ناموفق					موفق	ناموفق
۹۱.۵۷٪	۶۹۵	درست	۳۷۱	۲۵	۳۲۶	۹۲.۰۹٪	۱,۶۲۹	درست	۹۰۶	۷۷	۸۲۹
۸.۴۳٪	۶۴	نادرست	۳۸۸	۳۵۹	۲۹	۷.۹۱٪	۱۴۰	نادرست	۸۶۳	۸۰۰	۶۳
	۷۵۹	جمع	۷۵۹	۳۹۴	۳۶۵		۱,۷۶۹	جمع	۱۷۶۹	۸۷۷	۸۹۲

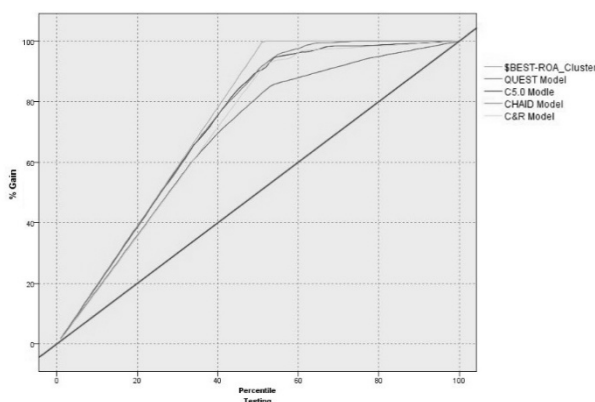
با توجه به جدول ۴ بر اساس معیار صحت، وضوح، نرخ مثبت نادرست، دقت و معیار F الگوریتم کاشف تعاملات خودکار کای دو در بیشترین مقدار قرار دارد و پس از آن به ترتیب الگوریتم خوشه‌بندی و رگرسیون، سی فایو و آماری موثر عاری از تعصب و سریع دارای بیشترین مقادیر هستند. با توجه به مقادیر نزدیک بدست آمده برای هر ۴ مدل به

طور کلی می‌توان گفت تمامی مدل‌ها از در بازه‌ی قابل قبولی قرار گرفته‌اند.

جدول ۴. نتایج حاصل از پیش بینی برای داده‌های آزمایشی

نوع مدل	صحت	حساسیت	وضوح	نرخ مثبت نادرست	نرخ منفی نادرست	دقت	نرخ خطا در طبقه بندی	معیار F
Quest	۸۲.۶۱٪	۸۲.۷۵٪	۸۲.۴۷٪	۱۸.۳۳٪	۱۶.۴۹٪	۸۱.۸۷٪	۱۷.۳۹٪	۸۲.۶۱٪
C۵.۰	۹۰.۳۸٪	۸۹.۷۶٪	۹۰.۹۸٪	۹.۴۳٪	۹.۱۹٪	۹۰.۴۹٪	۹.۶۲٪	۹۰.۳۴٪
CHAID	۹۱.۸۳٪	۸۸.۶۸٪	۹۴.۸۵٪	۵.۳۹٪	۱۰.۸۲٪	۹۴.۳۷٪	۸.۱۷٪	۹۱.۶۶٪
C&R	۹۱.۵۷٪	۹۰.۵۷٪	۹۲.۵۳٪	۷.۸۲٪	۹.۰۲٪	۹۲.۰۵٪	۸.۴۳٪	۹۱.۵۴٪

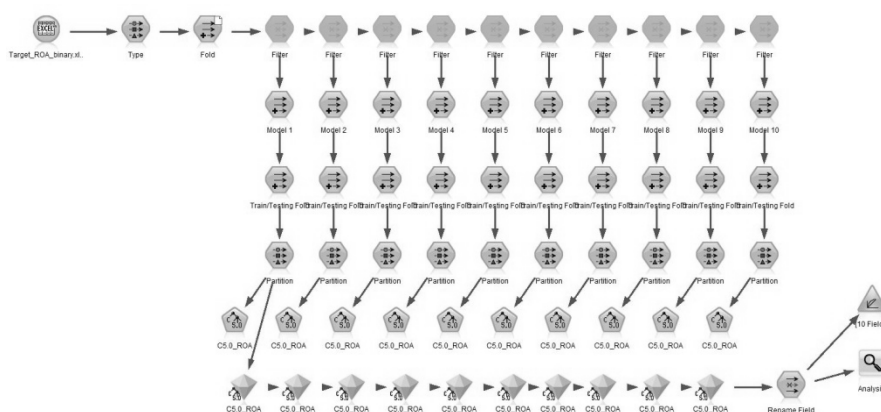
یکی دیگر از معیار ارزیابی استفاده از نمودار بهره است. بهره به صورت نسبتی از کل موفقیت‌ها است که در هر چندک اتفاق می‌افتد و به صورت (تعداد موفقیت‌ها در چندک‌ها/کل تعداد موفقیت‌ها) \* ۱۰۰٪ تعریف می‌شود (IBM, ۲۰۱۵). نتایج حاصل از نمودار بهره نیز همراستا با ارزیابی شاخصه‌هاست. با توجه به شکل ۱ درخت کاشف تعاملات خودکار کای دو در بیشتر صدک‌ها از عملکرد بهتری برخوردار بوده است.



شکل ۱. ارزیابی مجموعه داده‌های تست از طریق نمودار بهره

اعتبار سنجی متقابل داده‌های مورد بررسی را به  $K$  زیر مجموعه‌ی هم اندازه تقسیم می‌کند و یک زیر مجموعه را گروه آزمایش و دیگر زیر مجموعه‌ها را گروه یادگیری در نظر می‌گیرد. این کار تکرار می‌شود تا هر زیر مجموعه یک بار به عنوان گروه آزمایش و یک بار به عنوان گروه یادگیری انتخاب شود. در این پژوهش از ۱۰ تکرار برای

اعتبارسنجی استفاده شده است. شکل ۲ پیاده‌سازی اعتبارسنجی متقابل را برای الگوریتم سی فایو نشان می‌دهد. نتایج این گروه‌بندی به عنوان معیاری برای بررسی کیفیت قوانین حاصل و اعتبارسنجی مدل به کار می‌رود (Karimi & Sadeghi, 2014; Zhong, Dong, & Ohsuga 2001).



شکل ۲. مدل اعتبارسنجی متقابل برای  $k=10$

نتایج خلاصه شده حاصل از مدل اعتبارسنجی متقابل برای ۴ الگوریتم در جدول ۵ آورده شده است. با توجه به ارزیابی در این روش الگوریتم کاشف تعاملات خودکار کای دو دارای حداکثر صحت در داده‌های آموزش و آزمایشی (۹۶،۳۳، ۹۰،۷۶) می‌باشد. با توجه به میانگین متوسط صحت‌های ۱۰ مرتبه انجام ارزیابی الگوریتم سی فایو از عملکرد بهتری برخوردار است (۸۹،۳۷، ۸۷،۵۶) و نشان دهنده‌ی این موضوع است که دقت قوانین به دست آمده در این مدل از سایر مدل‌ها بیشتر است.

جدول ۵. نتایج اعتبارسنجی متقابل

نوع مدل	داده‌های آموزشی			داده‌های آزمایشی		
	حداقل صحت	حداکثر صحت	متوسط صحت	حداقل صحت	حداکثر صحت	متوسط صحت
CHAID	۸۰.۰۰٪	۹۶.۳۳٪	۸۶.۹۸٪	۸۱.۵۶٪	۹۰.۷۶٪	۸۵.۰۹٪
C5.0	۸۳.۶۴٪	۹۷.۳۷٪	۸۹.۳۷٪	۸۴.۱۰٪	۸۹.۷۳٪	۸۷.۵۶٪
Quest	۷۷.۰۸٪	۸۳.۷۹٪	۷۹.۵۷٪	۷۶.۰۴٪	۸۳.۸۲٪	۷۸.۸۴٪
C&R	۸۵.۲۵٪	۸۹.۳۴٪	۸۶.۹۷٪	۸۳.۲۳٪	۸۹.۴۰٪	۸۶.۷۸٪

نمودار اهمیت پیش بینی، نشان دهنده‌ی قدرت ارتباطی متغیرهای وابسته با متغیر هدف و مستقل از دیگر ورودی‌ها است.

به طور معمول در یک مدل تمرکز اصلی بر روی شاخص‌های پیش بینی کننده ای است که در مدل تاثیر بیشتری داشته باشند. این مقادیر با دقت و صحت مدل ارتباطی ندارند و تنها اهمیت آن‌ها را در ساخت مدل نشان می‌دهد و با نتایج داده‌های آزمایشی محاسبه می‌شوند. جدول ۶ میزان اهمیت متغیرهای مستقل را در چهار مدل مورد بررسی نشان می‌دهد.

جدول ۶. میزان اهمیت متغیرهای مستقل در الگوریتم‌های درخت تصمیم

C&R	QUEST	C5.0	CHAID	متغیرهای مستقل	C&R	QUEST	C5.0	CHAID	متغیرهای مستقل
		۰.۱۱	۰.۰۲	نسبت بدهی جاری به ارزش ویژه			۰.۰۱	۰.۰۵	نسبت جاری
		۰.۰۱		نسبت دارایی ثابت به ارزش ویژه					نسبت آبی
۰.۱۲		۰.۰۸		گردش دارایی ثابت			۰.۰۱		نسبت نقدینگی
۰.۱۵		۰.۱۲	۰.۰۸	گردش مجموع دارایی‌ها					نسبت دارایی جاری
				گردش سرمایه جاری			۰.۰۱		نسبت کفایت نقد
				نسبت کالا به سرمایه در گردش	۰.۰۵				نسبت گردش نقد
۰.۲۶		۰.۲۴	۰.۶۲	سود خالص به فروش			۰.۱۲	۰.۰۵	نسبت بدهی
				سود ناخالص به فروش	۰.۰۳	۰.۱۴			نسبت بدهی به ارزش ویژه
		۰.۰۳		سود عملیاتی به جمع درآمدها			۰.۰۲	۰.۰۱	نسبت پوشش بدهی
	۰.۲۴	۰.۰۱		سود ناویژه به فروش	۰.۲۳		۰.۲۲	۰.۱۳	نسبت بار مالی وام
		۰.۰۲	۰.۰۲	سود خالص به سود ناویژه		۰.۵۵			نسبت مالکانه

همان طور که اشاره شد مدل‌های C&R، CHAID و C5.0 از بیشترین صحت و دقت برخوردار بودند پس می‌توان گفت نتایج بدست آمده از آن‌ها قابل اطمینان تر می‌باشد. متغیرهای سود خالص به فروش، نسبت بار مالی وام و گردش مجموع دارایی‌ها به ترتیب نسبت‌های پر اهمیت در ساخت مدل‌ها می‌باشند.

اولین نسبت سود خالص به فروش نماینده‌ی متغیرهای سودآوری محسوب می‌شود. این نسبت نشان می‌دهد چند درصد از فروش را سود خالص تشکیل می‌دهد و تحت تاثیر عواملی همچون: حجم فروش، ترکیب بهای تمام شده کالای فروش رفته، سیاست قیمت گذاری، ساختار هزینه‌ها و کارایی واحد تولیدی قرار دارد.

نسبت بار مالی وام جز دسته‌ی متغیرهای بدهی است و توانایی شرکت در بازپرداخت اصل و بهره وام‌ها مورد سنجش قرار می‌دهد. توانایی بازپرداخت بدهی عبارتست از قدرت جوابگویی شرکت به بدهی‌های بلندمدت خود، در سررسید. تحلیل

این توانایی بر ساختار بلندمدت مالی و عملیاتی شرکت تاکید دارد. میزان بدهی‌های بلندمدت شرکت در ساختار سرمایه‌ی آن نیز در این تحلیل مورد توجه قرار می‌گیرد. در این بررسی‌ها سودآوری شرکت نیز باید مورد دقت قرار گیرد، چرا که تا شرکتی سودآور نباشد قادر به بازپرداخت بدهی‌های خود نخواهد بود. وقتی بدهی یک شرکت زیاد باشد مدیریت ناچار خواهد بود که برای تأمین منابع مالی اضافی به روش‌های دیگری و درصدد آن به صاحبان سرمایه (صدور سرمایه) متوسل شود و نیز تلاش نماید که سررسید بدهی‌ها را تمدید کند که بدون تردید هزینه‌ی تأمین سرمایه در هر یک از این حالات افزایش خواهد یافت. این نسبت به ما نشان می‌دهد که شرکت چند برابر مبالغی که بابت وام باید بپردازد در اختیار دارد.

گردش مجموع دارایی در دسته‌ی نسبت‌های کارایی قرار دارد. این نسبت کارایی شرکت را در استفاده از دارایی‌ها نشان می‌دهد به طوری که فروش سالانه ایجاد شده توسط هر ریال از دارایی‌ها را اندازه‌گیری می‌کند. در سال نرمال گردش جمع دارایی‌ها برای هر شرکت ۱ است که بدین معناست در هر سال ۱ ریال فروش به ازای هر ریال دارایی ایجاد شده است. در سال بد این به کمتر از ۱ و در سال خوب این نسبت به بالای ۱ افزایش می‌یابد. زمانی که شرکت به ظرفیت تولید خود برسد این نسبت حداکثر خواهد شد و در نتیجه هر چه این نسبت افزایش پیدا کند، بیانگر فعالیت بیشتر واحد تجاری است. گردش مجموع دارایی‌ها بیانگر ارتباط بین ارزش دارایی‌ها و حجم فعالیت‌ها نیز می‌باشد.

## بحث و نتیجه‌گیری

با توجه به تحقیقات انجام شده ارتباط بین نسبت‌های مالی و عملکرد شرکت یک مسئله مورد قبول است. اکثریت تحقیقات پیشین بر اساس روش‌های سنتی انجام گرفته است. نوآوری تحقیق حاضر برای پژوهش‌های حوزه‌ی نظام ارزیابی عملکرد را می‌توان استفاده از داده‌کاوی و الگوریتم‌های درخت تصمیم دانست که با توجه به استفاده از متغیرهای مورد بررسی بیشتر و واقعی‌تر، نتایج شفاف و مورد اطمینانی را در اختیار ذینفعان قرار می‌دهد. نهایتاً با مقایسه چهار مدل درخت تصمیم می‌توان بهترین مورد را شناسایی کرد و با استفاده از متغیرهای مستقل و وابسته آتی به پیش بینی عملکرد آینده و برنامه ریزی‌های مربوطه پرداخت و عملکرد شرکت را بهبود بخشید. با توجه به یافته‌های پژوهش و تجربیات به دست آمده در مقایسه ۴ درخت تصمیم در روش‌های ارزیابی تخمین نمونه آزمایشی، اعتبار سنجی متقابل و شاخصه‌های ارزیابی درخت تصمیم کاشف تعاملات خودکار کای دو دارای بیشترین صحت است. البته با توجه به نتایج خوب بدست آمده در مدل‌های دیگر (بالای ۸۰ درصد) می‌توان از تمامی مدل‌ها استفاده کرد. همچنین در نتایج داده‌کاوی حاضر ۳ نسبت سود خالص به فروش، نسبت بار مالی وام و گردش مجموع دارایی‌ها در میان ۲۹ نسبت ابتدایی از اهمیت بالاتری در پیش بینی عملکرد داشته و در این خصوص به شرکت‌ها و ذینفعان دیگر پیشنهاد می‌شود در ارزیابی عملکرد این سه نسبت را بیشتر مورد مطالعه قرار داده و همچنین با به کارگیری از مدل‌های بدست آمده می‌توانند با حداقل احتمال ۸۰ درصد عملکرد آتی شرکت را پیش بین کنند.

در تحقیق حاضر تلاش شده است با مرور ادبیات، مدل‌های ساخته شده جامع و کامل باشد و با در نظر گرفتن روش‌های مختلف ارزیابی از صحت کافی برخوردار باشند با این وجود به نظر می‌رسد امکان گسترش پژوهش حاضر همچنان وجود داشته باشد. لذا به همین منظور در راستای پژوهش‌های آتی می‌توان متغیرهای عملکرد دیگر را نیز مدل سازی کرده و نتایج بدست آمده از الگوریتم‌های حاضر را بررسی کرد. همچنین می‌توان از متغیرهای بازار نیز به جز متغیرهای صورت‌های مالی در مدل استفاده کرد.

## پی‌نوشت‌ها

- |  |  |
|--|--|
| 1. CHAID (Chi-squared Automatic Interaction Detection) | 4. C5.0  |
| 2. QUEST (Quick, Unbiased, Efficient Statistical Tree) | 5. CRISP-DM(Cross Industry Standard Process for Data Mining) |
| 3. Classification and Regression                       | 6. SPSS  |

## منابع

- هاشمی، سید عباس؛ حسینی، سید محسن؛ برعندان، سجاد. (۱۳۹۱). مقایسه محتوای فزاینده اطلاعاتی نسبت‌های نقدی و تعهدی برای ارزیابی عملکرد مالی شرکت‌ها با رویکردی داده‌کاوی. *پژوهش‌های حسابداری مالی*، ۴، ۶۳-۸۱.
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4), 589-609.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman & Hall/CRC.
- 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.
- 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
- Foroghi, D., & Monadjemi, A. (2011, June). Applying decision tree to predict bankruptcy. In *Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on (Vol. 4, pp. 165-169)*. IEEE.
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques (Vol. 12)*. Springer Science & Business Media.
- Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT press.
- Horrigan, J. O. (1965). Some empirical bases of financial ratio analysis. *The Accounting Review*, 40(3), 558-568.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied statistics*, 119-127.
- Karimi, T. & Sadeghi Moghadam, M. (2014). *Rough Set and Grey Set Theory: principles, applications and software's*. First edition. Tehran: Mehraban nashr. (In Persian)
- Karami, G. R., & Talaei, 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.
- Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques—A review. *European journal of operational research*, 180(1), 1-28.
- IBM, IBM SPSS Modeler 17 Modeling Nodes, 2015
- Miner, G., Nisbet, R. & Elder, J., 2009. *Handbook of Statistical Analysis and Data Mining Applications*. 1 ed. s.l.: Academic Press.
- Ross, S. A., Westerfield, R., & Jordan, B. D. (2008). *Fundamentals of corporate finance*. Tata McGraw-Hill Education.
- Wang, H., Jiang, Y., & Wang, H. (2009, November). Stock return prediction based on Bagging-decision tree. In *2009 IEEE International Conference on Grey Systems and Intelligent Services (GSIS 2009)* (pp. 1575-1580). IEEE.
- Yu, G., & Wenjuan, G. (2010, May). Decision tree method in financial analysis of listed logistics companies. *International conference on intelligent computation technology and automation*.
- Zhong, N., Dong, J., & Ohsuga, S., (2001). Using Rough Sets with Heuristics for Feature Selection. *Journal of Intelligent Information Systems*, 16 (3): 199–214.

..

## **Determining a model of financial performance of Tehran Stock Exchange companies based on data mining**

**Iman Raeesi Vanani**

**Ghasem Blue**

**Shohreh Zarkesh**

### **Abstract**

Determining the performance of companies using financial ratios has always been an attraction for many researchers. Identifying the financial factors that have the greatest impact on corporate performance is one of the important issues for decision makers. In this research, by employing data mining standards, 4 decision making trees including CHAID, QUEST, C&R and C5.0 have been implemented and compared with evaluation criteria. Also, effective financial ratios have been determined. For this purpose a sample consisting of 30 ratios in 673 listed firms in Tehran Stock Exchange between years 1390 to 1393 have been considered. The results show that CHAID has the most accurate and precision in comparison with 4 decision trees. Of course other models have high reliability (above 80 %) and can be used all of them. Three variables including “Net Profit to Sales”, “Loans interest rate” and “Total Asset Turnover” had the highest significance in predicting performance.

**Keywords:** Financial performance, Evaluation Model, Data mining, Decision tree.