فصلنامه علمی پژوهشی
دانش حسابداری و حسابرسی مدیریت
سال اول / شماره سوم/ پائیز ۱۳۹۱

داده‌کاوی و کشف تقلب‌های مالی

فریدون رهنمای رودپشتی
استاد دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران
rahnama@iau.ir

تاریخ دریافت: ۹۱/۲/۲۰ تاریخ پذیرش: ۹۱/۶/۳۰

چکیده

تقلب از پدیده‌های رایج و متداول در کسب و کار است و طبق بخش ۲۴ استانداردهای حسابرسی ایران عمل فریبکارانه یک یا چند نفر از مدیران،کارکنان یا اشخاص ثالث برای برخورداری از مزیتی ناروا عبارتست از هرگونه اقدام عمدی یا غیرقانونی. بنابراین پیشگیری با کشف تقلب‌های با اهمیت در صورتهای مالی، همواره کانون توجه سرمایه‌گذاران، قانونگذاران، استاندارد گذاران، مدیران و حسابرسان بوده است.

مقاله حاضر، اثربخشی تکنیک‌های داده‌کاوی در تشخیص رفتارهای متقلبانه شرکت‌هایی که صورت‌های مالی متقلبانه گزارش نموده‌اند را بررسی کرده تا عوامل مؤثر بر اینگونه رفتارها را شناسایی کرد. داده‌کاوی پل ارتباطی میان علم آمار، علم کامپیوتر، هوش مصنوعی، الگوشناسی، فراگیری ماشین و بارنمایی بصری داده‌ها می‌باشد که در چارچوب فرایندی، استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه داده‌های بزرگ امکان‌پذیر می‌گردد و استفاده از آن در تصمیم‌گیری در فعالیت‌های تجاری مهم نظیر ارتقاء کیفیت سودمندی اطلاعات از طریق شناسایی تقلب های مالی میسر می‌شود.

این پژوهش به روش شناخت تاریخی با بهره گیری از اسناد کتابخانه‌ایی و به پشتوانه پیشینه و تحقیقات محققان، شواهدی لازم جهت پاسخ به سؤالات تحقیق ارائه می‌کند. نتایج مطالعه نشان می‌دهد که اولاً، تکنیک‌های داده کاوی، در شناسایی در صورتهای مالی متقلبانه سودمند هستند. ثانیاً، داده کاوی، به‌عنوان کانون هدایت فکر در مدیریت کسب و کارها جهت کشف تقلب می‌تواند مورد توجه قرار گیرد.

واژه‌های کلیدی: داده‌کاوی، تقلب مالی، صورت‌های مالی متقلبانه.

Archive of SID
www.SID.ir

۱- مقدمه

امروز حسابرسان با مسؤلیتی که به‌طور فزاینده‌ای سخت‌تر می‌شود درگیر هستند و شواهد زیادی از حساب سازی در بکارگیری حسابداری وجود دارد و به‌صورت گسترده‌ای از آن استفاده می‌شود. که این امر، به‌عنوان پدیده‌ایی رایج و مخرب، کارایی، نقدینگی و امنیت بازراهای مالی جهانی، بازارهای سرمایه و بدهی به دلیل وجود تقلب در صورتهای مالی که عامل اساسی در بحران مالی اخیر و رکود اقتصادی جهانی می‌باشد، تهدید می‌شوند. مبتنی بر پژوهش Kirkos, Spathis and Manolopulos (2007) کوسیوارا سال ۲۰۰۲ را از نقطه نظر دفتر داری و ادعاها که دارای دستکاری‌هایی هست که همچنان ادامه دارد، سال ناگوار نامید (کوسیوارا ۲۰۰۴). بعضی گزارشات هزینه تقلب کسب و کارهای ایالات متحده آمریکا را بیش از ۴۰۰ بیلیون دلار در سال برآورد کرده است (وِلز ۱۹۹۷). سپادیز، دویومپوس و زوپویونیوس (۲۰۰۲) اعتقاد دارند که صورتهای مالی متقلبانه به‌صورت فزاینده‌ای به‌طور مکرر در طی سالیان اخیر در حال ازدیاد می‌باشد. مدیران متقلب را چنین می‌توان تعریف کرد: مشارکت در تقلب عمدی به وسیله مدیران که منجر به آسیب به سرمایه‌گذاران و اعتباردهندگان، به خاطر استفاده از صورتهای مالی به شدت گمراه کننده است. در طی فرایند حسابرسی، حسابرسان باید امکان تقلب مدیران را برآورد کنند. AICPA به صراحت تصدیق کرده که حسابرسان در برابر تقلب‌های کشف شده پاسخگو هستند (کولینا و سوتو، ۲۰۰۲). از این گذشته برای توجه به انتظارات از ایشان حسابرسان تکنیک‌های تحلیلی را برای مرور کارشان به خدمت می‌گیرند، چیزی که به آنها اجازه می‌دهد مانده حسابها را بدون آزمون مرتبط با تراکنش منحصر به فرد آن برآورد سازند. فراسر، هاردلی و لین (۱۹۹۷) تکنیک‌های تحلیلی احتمالی اجمالی را به‌صورت غیرکمی، نمونه کمّی و پیشرفته کمّی طبقه‌بندی نمودند. تکنیک‌های پیشرفته کمّی در بردارنده روش‌های پیچیده بکارگیری آمار و هوش مصنوعی مانند شبکه عصبی و تحلیل رگرسیون بود.

داده کاوی (DM[1]) یک فرآیند تکراری است که در آن، فرآیند به وسیله پوشش آن تعریف می‌شود، همچنین می‌تواند با روش‌های کاملاً خودکار یا دستی انجام شود. DM می‌تواند بیشترین سودمندی را در یک زمینه راهنمای طرح (سناریوی) تحلیلی اکتشافی داشته باشد در جایی که هیچ تصور قبلی در مورد چیزی که کسب خواهیم کرد تشکل نیافته باشد (کانتارزیک، ۲۰۰۲). بکارگیری تکنیک‌های داده کاوی برای طبقه‌بندی‌های مالی یکی از حوزه‌های پرثمر تحقیقاتی است، بسیاری از قوانین التزامی و واحدهای تحقیقاتی ویژه، افرادی را مامور شناخت فعالیت‌های متقلبانه در اختیار دارند که با موفقیت داده کاوی را بکار می برند. هرچند در مقابل سایر زمینه‌های آزموده شده مثل پیش‌بینی ورشکستگی و پریشانی مالی، تحقیقات کاربرد تکنیک‌های DM به منظور کشف تقلبات مدیریتی در حداقل قرار دارد (کالدرون وچِه ۲۰۰۲ با کوسکیوارا و مَنولوپوس، ۲۰۰۴).

در این مطالعه ما به دنبال آزمون عمیق اطلاعات عمومی مندرج در صورتهای مالی هستیم که در آنها صورتهای مالی متقلبانه[2] کشف شده به وسیله روش‌های داده کاوی، متغیر می‌باشد. هدف این تحقیق شناختن عواملی است که مورد استفاده حسابرسان برای ارزیابی احتمالی FFS قرار گیرد. اولین هدف اصلی عبارت است از کاربرد و سنجش بکارگیری روش‌های داده کاوی در حالات متفاوت بین مشاهدات تقلب و بدون تقلب است.

پژوهشهای پشتیبان برای این تحقیق متشکل از تحقیقات مرتبط برای کشف تقلبات مدیریتی با استفاده از آمار و هوش مصنوعی (AI[3]) و تکنیک‌های داده کاوی است که همگی داده‌های عادی و در دسترس صورتهای مالی را بکار گرفته‌اند. روش‌های AI این مزیت تئوریکالی را دارند که نمی‌تواند مفروضات اختیاری را در ورودیهای متغیر به آنها تحمیل نمود. هرچند نتایج روش‌های AI گزارش شده اندکی یا گهگاهی در شکل کلی اجرا همان نتایج روش‌های آماری است.

در این مطالعه نتایج پژوهش آزمون سه تکنیک داده کاوی برای کاربرد در کشف تقلبات مدیریتی مبتنی بر یافته‌های یک تحقیق کاربردی که در یونان آزموده شده است استفاده گردید. نظیر درخت تصمیم‌گیری، شبکه عصبی، شبکه گمانه زنی بیزی. سه روش در حالاتی که پیش‌بینی صحیحی دارند مقایسه می‌شوند. داده‌های ورودی اکثراً از نسبت‌های مالی مشتق شده از صورتهای مالی به‌عنوان نمونه ترازنامه و صورت سود و زیان تشکیل شده است. نمونه متشکل از ۷۶ شرکت تولیدی یونانی است. ارتباط بین متغیرهای ورودی و طبقه‌بندی خروجی‌ها به وسیله مدل‌های کسب و آشکار می‌شود.

Archive of SID
www.SID.ir

مقاله حاضر چهار بخش اصلی متمرکز شده است: بخش ۱ تحقیقات پیشین و مبانی نظری را مرور می‌کند. بخش ۲ آگاهی در مورد روش‌شناسی (متدولوژی) و ساختار سؤالات مورد استفاده در پژوهش را مهیا می‌کند. همچنین در این بخش، استقرار مدل‌ها و الگوها و روش‌ها نیز ارائه می‌شود، در بخش ۳، تحلیل نتایج یک تحقیق کاربردی در کشور یونان به‌عنوان مبنای یافته‌ها مدنظر قرار گرفت و در پایان یعنی بخش ۴ نتایج ملاحظه شده به همراه نتیجه‌گیری و مباحث پیرامون آن را ارائه می‌دهد.

۲- مبانی نظری و پیشینه تحقیقات

۲-۱- مبانی نظری

رهنمای رودپشتی و همتی (۱۳۸۹) در پژوهشی به تبیین مبانی علمی و تشریح مفاهیم مرتبط با موضوع پرداختند و معتقدند طبق تعریف انجمن بازرسان رسمی تقلب[۴] (ACFE)، تقلب عبارتست از هر عمل غیرقانونی که همراه با فریب، پنهانکاری و تخطی و نقض اعتماد باشد. این اعمال همراه با تهدید و یا فشار فیزیکی نمی‌باشد. تقلب توسط افراد یا سازمانها برای بدست آوردن پول، دارایی یا خدمات، اجتناب از پرداخت یا زیان خدمات و یا ایمن ساختن مزیت تجاری یا شخصی صورت می‌گیرد.

طبق تعریف انجمن حسابداران رسمی آمریکا[۵] (AICPA)، تقلب عبارتست از یک مفهوم وسیع غیر قانونی که براساس اینکه آیا عمل انجام شده از روی قصد و عمد بوده باشد یا خیر، از اشتباه متمایز شناخته می‌شود (گلدمن، کافمن، ۲۰۰۹).

طبق بند ۴ بخش ۲۴ استانداردهای حسابرسی ایران تلقب عبارتست از هر گونه اقدام عمدی یا فریبکارانه یک یا چند نفر از مدیران، کارکنان یا اشخاص ثالث برای برخورداری از یک مزیتی ناروا یا غیرقانونی، تقلب شامل وجود انگیزه برای ارتکاب تقلب و استفاده از فرصت جهت ارتکاب آن است.

تقلب در صورتهای مالی طبق تعریف رضایی و ریلی (۲۰۰۹) بدین شرح است: اظهار غلط عمدی و یا حذف مقادیر یا افشای لازم در صورتهای مالی به منظور فریب دادن استفاده‌کنندگان صورتهای مالی.

طبق بند ۶ بخش ۲۴ استانداردهای حسابرسی ایران، گزارشگری مالی متقلبانه با تحریف یا حذف مبالغ یا موارد افشا از صورتهای مالی به عمد و به منظور فریب استفاده‌کنندگان صورتهای مالی، سروکار دارد.

همچنین رهنمای رودپشتی و همتی (۱۳۸۹)، برخی از اصطلاحات متداول و بنیادی در حوزه تقلب نظیر مثلث تقلب، لوزی تقلب، انواع تقلب‌های داخلی و پرچم‌های قرمز آن، ابزارها و تکنیک‌های پایه کشف تقلب از جمله، ابزارهای ارزیابی ریسک تقلب، تکنیک‌های کشف تقلب با استفاده از نرم افزار حسابرسی (ACL) به‌عنوان تکنیک‌ها و ابزارهای حسابرسی مبتنی بر کامپیوتر[۶] به نقش و جایگاه پیاده سازی حاکمیت شرکتی به‌عنوان الگوی راهبری پیشرفته می‌پردازند و معتقدند، هفت وظیفه اساسی و اصلی حاکمیت شرکتی شامل سرپرستی و نظارت، مدیریت، رعایت مقررات، حسابرسی داخلی، حسابرسی مستقل، مشاور می‌باشد که رضایی (۲۰۰۷) عنوان می‌کند، چهار وظیفه از وظایف فوق شامل سرپرستی،مدیریت، حسابرسی داخلی و مستقل برای پیشگیری و کشف مؤثر تقلب در صورتهای مالی، ضروری می‌باشد.

در تبیین مبانی داده کاوی، یاس (۱۳۹۰)،عنوان می‌کند داده کاوی پل ارتباطی میان علم آمار، علم کامپیوتر، هوش مصنوعی، الگوشناسی، فراگیری ماشین و بازنمایی بصری داده می‌باشد و فرآیند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه داده‌های بزرگ و استفاده از آن در تصمیم‌گیری در فعالیت‌های تجاری مهم است. بنابراین، داده کاوی به‌صورت یک محصول قابل خریداری نمی‌باشد بلکه یک رشته علمی و فرآیندی است که بایستی به‌صورت یک پروژه پیاده‌سازی شود. اصولاً هرجایی که داده وجود داشته باشد، داده کاوی نیز معنا می یابد. این فرآیند، دارای مزایایی با اهمیت است از جمله:

۱) کمک به مدیران در تصمیم‌گیری
۲) یافتن مقالات خبری روزآمد
۳) کمک به کاربر برای جستجو در وب
۴) بهبود امر بازاریابی
۵) کشف کلاهبرداریها، تقلب و نابهنجارها.

Archive of SID

www.SID.ir

اساس داده کاوی بر مبنای سه فعالیت اصلی است که عبارتند از:

الف- حذف داده‌ها: داده‌های بی‌ارزش و عوامل بیرونی حذف می‌شوند.

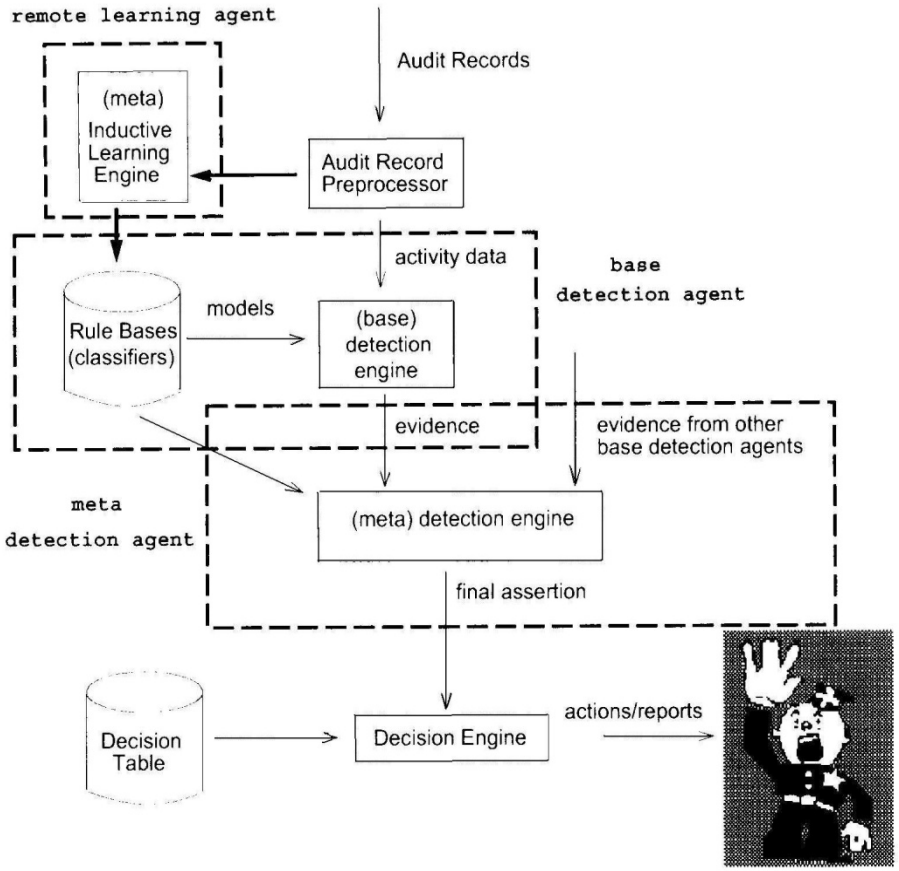
ب- فشرده‌سازی داده‌ها: این عمل بوسیله کدگذاری داده‌ها صورت می‌گیرد.

ج- کشف الگوها: الگوهای موجود در پایگاه داده‌ها از قبیل طبقه‌بندی، الگوهای زنجیری و ... کشف می‌شوند.

Lee & Stolf (1998)، چارچوبی تحت عنوان معماری داده‌کاوی به‌عنوان نظام پایه‌ایی کشف تقلب ارائه می‌دهد که در شکل شماره (۱) ارائه می‌شود.

Jans, Lybaert & Vanhoof [2007] از داده کاوی برای کشف تقلب، با رویکرد توسعه نظام کنترل داخلی نام می برد که به‌عنوان ماشین یادگیری تلقی گردید و کنترل داخلی به‌عنوان مکانیسم کشف تقلب معرفی شده است.

اهمیت و توسعه کاربرد داده کاوی در تجارت و اقتصاد و تصمیم‌گیری مدیریت موجب گردید که شرکتهای زیادی در این حوزه به فعالیت بپردازند و محصولات سخت افزاری و نرم‌افزارهای متعدد، متنوع و پیشرفته با رویکردها و ساختارهای متفاوتی به بازار عرضه کنند که حاصل پژوهش Abbott, Matkovsky & Elder در سال (۲۰۰۲)، منجر به شناسایی، معرفی و الگوریتم‌های بکار رفته آنها گردید که در جداول شماره ۱ الی ۶ ذیلاً ارائه شده است.



شکل ۱

www.SID.ir

Table 1: Data Mining Products Evaluated

Company	Product	Version
Integral Solutions, Ltd. (ISL) [5]	Clementine	4.0
Thinking Machines (TMC) [6]	Darwin	3.0.1
SAS Institute [7]	Enterprise Miner (EM)	Beta
IBM [8]	Intelligent Miner for Data (IM)	2.0
Unica Technologies Inc. [9]	Pattern Recognition Workbench (PRW)	2.5

Table 2: Software and hardware Supported

Product	Server	Client	Oracle Connect
Clementine	Solaris 2.X	X Windows	Server side ODBC
Darwin	Solaris 2.X	Windows NT	Server side ODBC
Enterprise Miner	Solaris 2.X[1]	Windows NT	SAS Connect>for Oracle
Intelligent Miner	IBM AIX	Windows NT	IBM Data Joiner>
PRW	Data only	Windows NT	Client side ODBC

1 most testing performed on a standalone Windows NT version.

Table 3: Algorithms Implemented

Algorithm	IBM	ISL	SAS	TMC	Unica
Decision Trees	✓	✓	✓	✓	
Neural Networks	✓	✓	✓	✓	✓
Regression	[1]	✓	✓		✓
Radial Basis Functions	✓[2]				✓
Nearest Neighbor			✓		
Nearest Mean Kohonen Self-Organizing Maps		✓	✓		
Clustering	✓	✓			✓
Association Rules	✓	✓			

1 accessed in data analysis only 2 estimation only (not for classification)

Table 4: Options for Decision Trees

Algorithm	IBM	ISL	SAS	TMC
Handles Real-Valued Data	✓	✓	✓	✓
Costs for Misclassification		✓	✓	✓
Assign Priors to Classes			✓	✓
Costs for Fields	✓			
Multiple Splits			✓	
Advanced Pruning Options		✓		
Graphical Display of Trees	✓	✓	✓	

Table 5: Options for Neural Networks

Algorithm	IBM	ISL	SAS	TMC	Unica
Automatic Architecture Selection	✓	✓			
Advanced learning Algorithms		✓	✓	✓	✓
Assign Priors to Classes					✓
Costs for Misclassification					
Cross-Validation Stopping Rule		✓	✓	✓	✓

Table 6: Ease of Use Comparison

Category	IBM	ISL	SAS	TMC	Unica
Data Load and Manipulation	3.1	3.7	3.7	3.1	3.9
Model Bulding	3.1	4.6	3.9	3.2	4.8
Model understanding	3.2	4.2	2.6	3.8	3.8
Technical Support	3.0	4.0	2.8	3.2	4.7
Overlall Usability	3.1	4.1	3.1	3.4	4.2

Archive of SID

www.SID.ir

۲-۲- تحقیقات پیشین

مبتنی بر پژوهش Kirkos, Spathis & Manolopoulos (2007) در ۱۹۹۷، هیات استانداردهای حسابرسی استاندارد حسابرسی (SAS) شماره ۸۲ را صادر کرد که: رسیدگی به تقلب در حسابرسی صورت‌های مالی است. این استاندارد حسابرسان را ملزم به رسیدگی به ریسک تقلب در رسیدگی حسابرس و تشویق حسابرسان برای رسیدگی به دو موضوع سیستم کنترل داخلی و کنترل رویکرد گرایشات مدیریت در زمان ارزیابی را در بر دارد. عوامل ریسک یا «پرچم قرمز» مرتبط با گزارش صورت‌های مالی متقلبانه ممکن است در سه گروه به این شرح طبقه‌بندی شوند (SAS شماره ۸۲):

الف- خصوصیات مدیریت و نفوذ به محیط کنترلی؛

ب- خصوصیات صنعت؛

ج- خصوصیات اجرایی و پایداری مالی.

کمیته بین‌المللی مسایل حسابرسی (IAPC) از فدراسیون بین‌المللی حسابداران استاندارد بین‌المللی (ISA) شماره ۲۴۰ را صادر نموده است. این استاندارد انتظار رسیدگی به وسیله حسابرسان به ریسک تقلب و اشتباه را دارد و به وضوح استدلال‌هایی راجع به محدودیت‌های ذاتی بر توانایی حسابرس برای کشف اشتباه و خطا به ویژه تقلب مدیریت را بر شمرده است. علاوه بر این، آن تقویت کننده تمایز بین تقلب مدیریت و کارکنان و علل مفهومی تقلبات مالی است که می‌بایست با جزئیات گزارش شود.

کشف تقلب مدیریتی وقتی در فرآیند معمولی حسابرسی استفاده می‌شود وظیفه سختی است (پورتر و کامرون، ۱۹۸۷؛ کودیرِ، ۱۹۹۹). نخست اینکه آگاهی مفهومی کمی از ویژگی‌های تقلب مدیریتی وجود دارد. دوم، کسب آن کمیاب است، اکثر حسابرسان فاقد تجربه لازم در کشف آن هستند. در پایان، مدیران کنکاشانه تلاش می‌کنند حسابرسان را فریب دهند (فانینگ و کوگِر، ۱۹۹۸). برای چنین مدیرانی که محدودیت‌های هر حسابرسی را درک کرده‌اند فرایند حسابرسی استاندارد ممکن است ناکافی از اب درآید. این محدودیت‌ها اشاره به نیاز فرایند تحلیل‌های اضافه برای کشف موثر تقلب مدیریت دارد. این همچنین رشد توجه به سامانه ارزیابی موارد غیرعادی با یک وضعیت حرفه‌ای در مورد کشف تقلب را برجسته کرده است. همان جایی که اکثر تقلب‌های مهم در سطوح بالای سازمان، سازمان‌دهی می‌شود، جایی که کنترل‌ها و سیستم‌ها کمترین شیوع و اثر را دارند (کولینا وسوتون، ۲۰۰۲).

مطالعات جدید متوجه ساخت مدل‌هایی شده‌اند که تقلب پیش آمده مدیریت را پیش‌بینی کند. نتایج حاصل از تحلیل رگرسیون لاجیت ۷۵ شرکت متقلب و ۷۵ شرکت بدون تقلب اشاره به این موضوع داشته است که هیات مدیره شرکت‌های بدون تقلب به‌طور معنی‌داری با درصد بیشتری از اعضائ خارج از سازمان نسبت به شرکت‌های متقلب تشکیل شده است (باسلی، ۱۹۹۶). هاتسن، مکدونالد، میژر و بل (۱۹۹۶) یک مدل واکنشی کیفی تعمیم یافته مفید برای پیش‌بینی تقلب مدیران براساس دسته‌ای از داده‌های توسعه یافته به وسیله یک شرکت حسابرسی عمومی بین‌المللی را بکار بردند. این مدل در بردارنده تکنیک‌های پرابیت و لاجیت می‌باشد. نتایج اشاره به ظرفیت خوب پیش‌بینی دو گروه فرضیات مخارج متقارن و نامتقارن داشت.

پیشینه تحقیقات در خصوص کاربردی داده کاوی در پیشگیری از سوء استفاده‌های مالی و تقلب:

ردیف	محقق	سال	تکنیک داده کاوی	نتایج
۱	اینیگ، جونز و لوبک	۱۹۹۷	سیستم خبره	به شکل کاملاً مشخصی، بهتر می‌توان، وضعیت سرتاسر مراحل متغیر مخاطرات تقلب مدیریت را بررسی نمود
۲	گرین و چوی	۱۹۹۷	شبکه عصبی	به حسابرسان رهنمود مناسبی برای کشف تقلب ارائه می‌دهد
۳	فانینگ و کوگِر	۱۹۹۷	شبکه عصبی	نتایج بهتری در مقایسه با رگرسیون لاجستیک و مدل‌های آماری استاندارد ارائه داده است
۴	سامرز و سوعِنی	۱۹۹۸	مدل منطقی آبشاری	در تقلب‌های مشاهده شده، کارکنان داخلی، سهام خودشان را از شرکت در سطوح بالای فعالیت‌های فروش کاهش می‌دهند

Archive of SID
www.SID.ir

ردیف	محقق	سال	تکنیک داده کاوی	نتایج
۵	Chan و همکاران او	۱۹۹۹	داده کاوی توزیعی	امکان کشف تقلب کارتهای اعتباری جهت مدیریت ریسک فراهم می‌شود
۶	«سرولو» و «سرولو»	۱۹۹۹	شبکه عصبی	تقلب در گزارشگری مالی را پیش‌بینی نمود
۷	بل و کارسلو	۲۰۰۰	رگرسیون لاجستیک	احتمال گزارش مالی متقلبانه برای ادعای حسابرس،در حالتی که ریسک تقلب وجود دارد یا ندارد را افزایش می‌دهد
۸	اسپادیس	۲۰۰۲	رگرسیون لجستیک	احتمال بیشتر در تقلب و دستکاری در صورتهای مالی پیش‌بینی شد
۹	اسپادیس و سایرین	۲۰۰۲	داده کاوی توزیعی (UTADIS)	عملکردی بهتر در کشف تقلب در صورتهای نسبت به سایر روش‌های آماری (رگرسیون لجستیک و تجزیه و تحلیل آماری) دارد
۱۰	«گو» و «لو»	۲۰۰۴	درخت تصمیم‌گیری	مشکلات مخفی موجود در صورتهای مالی از طریق نسبت‌های مالی کشف گردید
۱۱	لین، هوانگو وبکر	۲۰۰۸	شبکه عصبی و عصبی فازی	هر دو مدل قادر به ارزیابی و طبقه‌بندی گزارش‌های مالی شرکت‌ها به متقلبانه و غیرمتقلبانه هستند
۱۲	لیو	۲۰۰۸	رگرسیون لجستیک شبکه عصبی و درخت تصمیم‌گیری	با توجه به نتایج بدست آمده رگرسیون لجستیک، و شبکه عصبی قادر به طبقه‌بندی درست شرکت‌های متقلب می‌باشد. ولی صحت و درستی طبقه‌بندی براساس الگوریتم درخت تصمیم‌گیری پایین است و رگرسیون لجستیک یک روش مناسب برای کشف گزارشگری مالی متقلبانه و پیش‌بینی شکست تجاری شرکت‌ها می‌باشد
۱۳	گلنی و «یاداو»	۲۰۱۰	مدل کشف تقلب محاسبه‌ایی (مدل متنی)	این مدل بر روی تحلیل‌های موجود در فایل K-۱۰ بورس اوراق بهادار امریکا، آزمون شد که قادر است بین گزارش‌های همراه با تقلب و بدون تقلب تفکیک قایل شود
۱۴	«زو» و «کاپو»	۲۰۱۰	رگرسیون لجستیک، درخت تصمیم‌گیری، شبکه عصبی و شبکه بیزیز	مدل‌ها قادر به کشف موثرتر تقلب در صورتهای مالی هستند
۱۵	حسینی‌نسب و همکاران	۱۳۹۰	داده کاوی روش Apriori و ترکیب اطلاعات روش Dempster-shaper و OWA	احتمال رخداد تقلب در نسخه‌های دارویی مربوط به یک فرد برای هر یک از زیرمجموعه داروهای مندرج در نسخ و توسط روش‌ها پیش‌بینی می‌شود
۱۶	Kirkos و همکاران	۲۰۰۷	درخت تصمیم‌گیری و شبکه عصبی و شبکه‌های گمانه‌زنی بیزی	نتایج در ادامه ارائه می‌شود

۳- روش‌شناسی و سؤالات اصلی پژوهش

روش پژوهش حاضر، روش شناخت تاریخی مبتنی بر پیشینه تحقیقات با رویکرد فلسفی جهت ارزیابی توان تبیین داده کاوی در کشف تقلب صورتهای مالی است که از نتیجه تحقیقی در یونان که توسط Kirkos, Spathis & Manolopoulos در سال ۲۰۰۷ صورت گرفت به‌عنوان مطالعه موردی مدنظر قرار گرفت و مبانی بکار گرفته در پژوهش محققان فوق مبتنی بر تحقیقات دیگران بوده است که عبارتند از:

۳-۱- داده‌ها

نمونه تحقیق متشکل از داده‌های ۷۶ شرکت تولیدی یونانی (بدون اینکه شرکت‌های مالی در آن باشد) می‌باشد. حسابرسان تمام ۷۶ شرکت نمونه را بررسی کرده‌اند. در ۳۸ شرکت از این شرکت‌ها نشانه یا دلیلی از

Archive of SID
www.SID.ir

وجـود صـورت‌هـای مـالی متقلبانـه اعـلام شـده اسـت. طبقه‌بندی صورت‌هـای مـالی بـه‌عنـوان خطـا مبنـی بـر معیارهـای معـین بـه ایـن شـرح اسـت: نتیجـه گـزارش حسابرسی راجع به تردیـدهای جـدی راجـع بـه درسـتی حسابها، اسناد مالیاتی راجع به عدم تراضی مالیاتی جـدی که از دگرگونی شدید ترازنامه و صورت سود و زیان سالانه شرکت در بکـارگیری قـوانین یونـان در مـورد رد خـالص ارزش آنها می‌باشد کـه عاقبـت چنـین شـرکت هـایی در بورس آتن قرار گرفتند که در زمره شرکتهای تحت نظر و «معلق در حال مذاکره» برای مشخص شدن دلایل دخیل در تحریف داده‌های مالی شرکت‌ها و ارسـال پرونـده آنهـا جهت گذراندان، طی فرایند قضایی با اتهام FFS یا جـرایم شدید مالیاتی است.

۳۸ شرکت دارای تقلـب بـا ۳۸ شـرکت بـدون تقلـب تطبیق داده شد. شرکت هایی که بـه‌عنـوان شـرکت‌هـای بدون FFS مشخص شده‌اند براساس نبـود (فقـدان) هـیچ اشـاره یـا گـواهی رد بـاب گزارشـی از FFS در گزارشـات حسابرسی در امور مالی و پایگـاه داده مالیـاتی در بـورس سهام آتن مـی‌باشـد. ایـن مسـیر تضـمین نمـی‌کنـد کـه صورتهای مالی این شرکت‌ها بدون تحریـف یـا رفتارهـای FFS باشند که در آینده شناسایی شود. ایـن تنهـا FFS پیدانشده در یک بررسی مرتبط مبسوط را تضـمین کـرده اسـت. تمـام متغییرهـای مـورد اسـتفاده از نمونـه‌هـا از صورتهای مالی رسمی مانند ترازنامه‌ها و صورتهای سـود و زیـان اسـتخراج شـده اسـت. ایـن بـدان معنـا اسـت کـه سودمندی این مطالعه به واسطه این واقعیات کـه تنهـا از داده‌های شرکت‌های یونانی استفاده شـده اسـت، محـدود نمی‌شود.

۲-۳- متغیرهای پژوهش

گزینش متغییرهـا بـا اسـتفاده از معرفـی داده‌هـای خـط دهنده‌ی با مشارکت مبتنـی بـر کارهـای تحقیقـاتی پیشـین مرتبط با موضوعات FFS بوده است. مانند کار انجـام شـده بـه وسـیله‌ی اسـپادیس (۲۰۰۲)، اسـپادیس و سـایر همکـاران (۲۰۰۲)؛ فانینگ و کوگر (۱۹۹۸)؛ پرسونس (۱۹۹۵)؛ استیک (۱۹۹۵)؛ فـروز، پـارک و پاسـتنا (۱۹۹۱)؛ لئوبیـک، اینینـگ، ویتینـگ هـام (۱۹۸۹) و کنـی و مـک دانـل (۱۹۸۹) کـه در بردارنده پیشنهادهایی با اشاره به FFS می‌باشد.

پریشانی مـالی ممکـن اسـت یـک مشـوق بـرای تلقبـات مـدیریتی باشـد (فانینـگ و کـوگر، ۱۹۹۸؛ اسـتیک، ۱۹۹۱؛ لیوبیک و سایر همکاران، ۱۹۹۸؛ کینی و مک دانل، ۱۹۸۹). به منظور تطبیق نمودن یک نسـبت در بـاب پریشـانی مـالی مـا امتیاز Z آلتمن را که مشهور است را برای برآورد پریشانی مالی بکار گرفتیم (آلتمن، ۱۹۶۸). از این رو که بسیاری از تحقیقات در مطالعات مرتبط با پیش‌بینی ورشکستگی و پریشانی مـالی به‌طور گسترده امتیاز Z را استفاده کرده‌اند.

پیرسوس (۱۹۹۵) معتقد بود که یک پرسـش مطـرح این است که آیا یک سـاختار بـدهی سـنگین در گـزارش مالی متقلبانه دخالت دارد. یـک سـاختار بـدهی سـنگین ممکن است احتمال FFS را افزایش دهد بدین صورت که این موضوع بـا ریسـک حقـوق مالکـان و مـدیریت بـدهی توسـط مالکـان مـرتبط اسـت. مـدیران ممکـن اسـت بدهی‌هایی را که براساس قراردادهای بدهی در صـورتهای مالی در حال سررسید شدن می‌باشد را دسـتکاری کننـد. این موضوع اشاره به این دارد که سطوح بالاتری از بـدهی ممکن است احتمـال FFS را افـزایش دهـد. مـا آن را بـه وسـیله اسـتفاده از لگـاریتم کـل بـدهی (LOGDEBT)، نسبت بدهی بـه ارزش ویـژه (DEBTEQ) و نسـبت کـل بدهی به کل دارایی‌ها (TDTA) اندازه گرفته‌ایم.

انگیزه دیگر تلقب مدیریتی نیاز به رشد پایـدار اسـت. شرکتهای ناتوان در سیدن به نتایج مشابه عملکرد گذشته ممکن است درگیر فعالیتهای متقلبانه بـرای حفـظ رونـد پیشین خود شـوند (اسـتیک، آلبرچـت و بـراون، ۱۹۹۱). شرکت‌هایی کـه رشـد سـریعی داشـته‌انـد ممکـن اسـت فرایندی را تحت نظر قرار دهند که قابلیت نظارت مناسب را تـدارک مـی‌بینـد و آن را پشـت سـر گذاشـته باشـند (فانینـگ و کـوگر، ۱۹۹۸). بـرای انـدازه گیـری رشـد، از نسبت رشد فروش (SALGRTH) استفاده می کنیم.

حسابرسی معدودی از حسابهایی که استفاده از برآورد در آنها مجاز است، مشکل می‌باشد در نتیجه آنها مسـتعد تحریف‌هـای متقلبانـه هسـتند. حسـاب هـای دریـافتنی، موجودی و فروش‌ها در این گروه قرار می‌گیرند. پیرسونس (۱۹۹۵)؛ اسـتیک (۱۹۹۱) و فـروز و همکـاران (۱۹۹۱)

Archive of SID
www.SID.ir

معتقدند که مدیریت ممکن است در حسابهای دریافتنی دستکاری کند. فعالیت‌های متقلبانه ثبت فروش قبل از اینکه تحقق یابد ممکن است حسابهای دریافتنی را اضافه‌تر نشان دهد (فانینگ و کوگر، ۱۹۹۸). در این تحقیق، حسابهای دریافتنی را با استفاده از نسبت حسابهای دریافتنی / فروش (RECSAL) و نسبت حسابهای دریافتنی / حسابهای دریافتنی برای دو سال پیاپی (RETREND) و متغییر رقومی (دیجیتال) REC11 که اشاره به تغییرات ۱۰٪ دارد مدنظر قرار گرفت.

بسیاری از تحقیقات به این موضوع اشاره دارد که ممکن است مدیریت موجودی‌ها را دستکاری کند (استیک ۱۹۹۱؛ پیرسون ۱۹۹۵؛ اس چلیت ۲۰۰۲). گزارش موجودیها با بهای تمام شده کمتر و ثبت موجودیهای راکد بعضی از تدابیر شناخته شده است. ما موجودی‌ها را به وسیله استفاده از نسبت موجودی / فروش (INVSAL) و موجودی به مجموع دارایی‌ها (INVTA) بررسی می‌کنیم. حاشیه سودناخالص نیز مستعد دستکاری است. شرکتها ممکن است تطابق فروش‌هایشان را با بهای تمام شده کالای فروش رفته مربوط انجام ندهند. در نتیجه حاشیه سود ناخالص، خالص درآمد و استحکام ترازنامه افزایش یابد (فانینگ و کوگر ۱۹۹۸). ما حاشیه سود ناخالص را با استفاده از نسبت فروش منهای حاشیه سودناخالص (COSAL) و نسبت سود ناخالص / مجموع دارایی‌ها (GPTA) و نسبت حاشیه سود ناخالص / حاشیه سود ناخالص دو سال پیاپی (GMTREND) و متغییر دو تایی باینری GM11 که اشاره به ۱۰٪ افزایش نسبت به ارزش سال پیش دارد، آزمون می‌کنیم.

اسپادیس (۲۰۰۲) در یک مطالعه رگرسیون لاجستیک برای پیش‌بینی FFS معتقد بود سود خالص / مجموع دارایی‌ها (NPTA) و سرمایه در گردش / مجموع دارایی‌ها (WCTA) نسبت‌های معنی‌داری هستند. به علاوه اسپادیس و همکارانش (۲۰۰۲) به این موضوع اشاره داشتند که همچنین نسبت سود خالص / فروش (NPSAL) بیشترین معنی داری را در خود دارد. در این مطالعه، ارتباط بعضی از «پرچم‌های قرمز» صورتهای مالی با قابلیت پیش‌بینی FFS آنها آزموده شد. این نسبت‌ها عبارتند از: لگاریتم مجموع دارایی‌ها (LTA)، سرمایه در گردش (WCAP)، نسبت تاسیسات و تجهیزات (net fixed assets) به مجموع دارایی‌ها (NFATA) و فروش به مجموع دارایی‌ها (SALTA) و دارایی‌های جاری / بدهی‌های جاری (CACL) و خالص درآمد / دارایی‌های ثابت (NIFA) و موجودی نقد / مجموع دارایی‌ها (CASHTA) و دارایی‌های آنی / بدهی‌های جاری (QACL)، سود قبل از بهره و مالیات (EBIT) و بدهی‌های بلند مدت / مجموع دارایی‌ها (LTDTA).

در مطالعه Kirkos و همکاران او ۲۷ نسبت مالی را گردآوری گردید. در کوششی برای کاهش ابعاد، آنالیز واریانس یک طرفه ANOVA را به منظور آزمون اینکه آیا اختلاف دو گروه برای هر متغییر معنی‌دار است، اجرا گردید. اگر اختلاف معنی‌دار نبود (P-value بالا) مغییر به‌عنوان اطلاعات غیرمفید تلقی می‌شود. جدول ۷ میانگین، انحراف استاندارد و F-value و P-value را برای هر متغییر نشان می‌دهد.

همانطور که در جدول ۷ می‌بینید، ۱۰متغییر P-value کم ($P \leq ۰/۰۵$) را نمایش می‌دهد. این متغییرها به‌عنوان متغیرهای دخالت کننده در داده‌های خط دهنده انتخاب می‌شوند و از بقیه متغیرها تفکیک می‌گردد. متغیرهای انتخاب شده، SALTA، COSAL، EBIT، WCAP، ZSCORE، TDTA، NDTA، WCTA، GPTA، DEBTEQ هستند.

۳-۳- روش‌ها و الگوهای آزمون

شناسایی صورت‌های مالی متقلبانه را می‌توان به‌عنوان نوعی طبقه‌بندی مشکل در نظر داشت. طبقه‌بندی یک رویه دو مرحله‌ای دارد. گام اول، یک مدل که با استفاده از نمونه‌های آموزشی، آموزش می‌بینند. نمونه با چندین متغییر (به شکل سطرها) و توزیع آنها (در ستون‌ها) سازمان دهی می‌شود. یکی از توزیع‌ها و برچسب توزیع دربردارنده ارزش معرف از پیش تعیین شده طبقه برای هر کدام از چندین متغییر متعلق به آن متغییر می‌باشد. این مرحله به‌عنوان آموزش هدایت شده شناخته می‌شود.

Archive of SID
www.SID.ir

در گام دوم مدل تلاشی می‌شود موضوعاتی را بـا اسـتفاده از اعتبار نمونه‌ها بر پا کرد که به نمونه‌های آزمایشی تعلق ندارند.

فرایندهای داده کاوی روش‌های چندین دسـته بنـدی است که از رشته‌هایی از آمار و هوش مصنوعی بـه دسـت آمده است. سه روش که از شهرت خوب در قابلیت طبقـه بندی بهره‌مندند در ایـن مطالعـه نیـز از تحقیقـات دیگـر محققان نیز استفاده گردید. این روش‌ها عبارتند از درخت تصمیم‌گیری، شبکه‌های عصبی و شـبکه‌هـای گمانـه‌زنـی بیزی می‌باشند.

جدول شماره ۷: نتایج آزمون

Variables	Mean FFS	SD FFS	Mean non-FFS	SD non-FFS	F	p-value
۱- بدهی به ارزش ویژه (DEBTEQ)	۲/۷۰۶	۳/۵۳۱	۱/۰۷۵	۰/۹۳۷	۷/۵۶	۰/۰۰۷
۲- فروش به مجموع دارایی‌ها (SALTA)	۰/۶۹۹	۰/۴۱۶	۱/۰۵۵	۰/۵۷۷	۹/۵۳	۰/۰۰۳
۳- سود خالص به فروش (NPSAL)	-۰/۴۵۹	۲/۴۳۴	۰/۰۶۷	۰/۱۵۹	۱/۷۷	۰/۱۸۸
۴-حسابهای دریافتنی به فروش (RECSAL)	۱/۷۵۵	۵/۸۹۷	۰/۴۵۶	۰/۳۴۹	۱/۸۴	۰/۱۷۹
۵- تاسیسات و تجهیزات به کل دارایی (NFATA)	۰/۳۲۰	۰/۱۹۳	۰/۲۷۶	۰/۱۴۰	۱/۳۱	۰/۲۵۷
۶- حسابهای دریافتنی به حسابهای دریافتنی برای دو سال پیاپی (RETRENET)	۱/۲۷۳	۰/۴۹۱	۱/۷۳۳	۳/۰۲۸	۰/۸۶	۰/۳۵۸
۷-متغیر رقومی (دیجیتال) نسبت ۶ به میزان ۱۰٪ (REC11)	۰/۶۰۵	۰/۴۹۵	۰/۷۱۱	۰/۴۶۰	۰/۹۲	۰/۳۴۰
۸- حاشیه سود ناخالص به سود ناخالص برای دو سال پیاپی (GMTREND)	۲/۵۶۴	۹/۰۱۴	۱۸/۷۱۵	۱۰۸/۴۸۴	۰/۸۴	۰/۳۶۳
۹- متغیر دوتایی باینری نسبت ۸ به میزان ۱۰٪ (GM11)	۰/۳۹۵	۰/۴۹۵	۰/۵۷۹	۰/۵۰۰	۲/۶۰	۰/۱۱۱
۱۰- رشد فروشی (SALGRTH)	۱/۱۶۳	۰/۷۰۵	۴۱۵/۶۱۶	۲۵۵۴/۴۶۴	۱/۰۰	۰/۳۲۰
۱۱- موجودی به فروشی (INVSAL)	۰/۳۵۹	۰/۵۶۵	۰/۱۷۹	۰/۱۵۹	۲/۷۰	۰/۱۰۵
۱۲- موجودی به کل دارایی‌ها (INVTA)	۰/۱۹۱	۰/۱۶۰	۰/۱۷۸	۰/۱۴۲	۰/۱۳	۰/۷۳۲
۱۳- موجودی نقد به کل دارایی‌ها (CASHTA)	۰/۰۷۱	۰/۱۲۲	۰/۰۹۶	۰/۱۱۴	۰/۸۶	۰/۳۵۶
۱۴- لگاریتم مجموع دارایی‌ها (LTA)	۸/۴۵۱	۰/۹۹۶	۸/۴۷۳	۰/۲۸۸	۰/۰۱	۰/۹۳۳
۱۵- لگاریتم مجموع بدهی‌ها (LOGDEBT)	۷/۸۷۰	۱/۲۳۳	۷/۵۱۲	۱/۴۱۱	۱/۳۹	۰/۲۴۳
۱۶- فروش منهای حاشیه سود ناخالص (COSAL)	۳۱۹۶/۶۰۰	۳۰۶۶/۱۸۸	۵۸۸۹/۷۴۲	۷۷۳۴/۲۸۵	۳/۹۸	۰/۰۵۰
۱۷- سود قبل از بهره و مالیات (EBIT)	-۳۳/۶۲۴	۱۳۹۲/۹۳۰	۱۰۹۳/۲۵۳	۲۸۷۵/۹۲۹	۴/۷۳	۰/۰۳۳
۱۸- سرمایه درگردش خالص (WCAP)	۳۶۹/۱۷۱	۲۱۶۶/۶۸۵	۲۲۸۰/۷۳۷	۴۲۷۹/۶۸۵	۶/۰۳	۰/۰۱۶
۱۹- Z آلتمن (ZSCORE)	۰/۷۷۸	۰/۹۳۶	۱/۹۹۰	۰/۷۳۰	۳۹/۵۹	۰/۰۰۰
۲۰- خالص درآمد به دارایی ثابت (NIFA)	-۰/۵۳۱	۱/۹۹۶	-۱/۵۵۴	۱۱/۶۳۵	۰/۲۹	۰/۵۹۵
۲۱- کل بدهی به کل دارایی (TDTA)	۰/۶۲۹	۰/۲۴۳	۰/۴۳۷	۰/۱۹۶	۱۴/۳۱	۰/۰۰۰
۲۲- سود خالص به کل دارایی (NPTA)	-۰/۰۲۱	۰/۰۹۵	۰/۰۷۴	۰/۰۶۵	۲۶/۱۱	۰/۰۰۰
۲۳- دارایی جاری به بدهی جاری (CACL)	۲/۱۱۳	۶/۰۹۵	۲/۳۰۲	۲/۲۶۷	۰/۰۳	۰/۸۵۹
۲۴- سرمایه در گردش به کل دارایی (WCTA)	۰/۰۵۴	۰/۲۳۹	۰/۲۵۳	۰/۲۰۵	۱۵/۱۵	۰/۰۰۰
۲۵- دارایی آنی به بدهی جاری (QACL)	۱/۱۶۳	۲/۲۴۵	۱/۷۷۲	۱/۸۰۴	۱/۷۰	۰/۱۹۶
۲۶- سود ناخالص به کل دارایی‌ها (GPTA)	۰/۱۴۴	۰/۱۲۱	۰/۲۷۴	۰/۱۴۰	۱۸/۷۸	۰/۰۰۰
۲۷- بدهی بلندمدت به کل دارایی (LTDTA)	۰/۰۶۵	۰/۱۳۸	۰/۰۴۱	۰/۰۷۷	۰/۸۷	۰/۳۵۳

Archive of SID
www.SID.ir

۳-۳-۱- درخت تصمیم‌گیری

یک درخت تصمیم‌گیری (DT) یک ساختار درختی است که هر گره آن یک آزمون برای یک صفت را نشان می‌دهد و هر انشعاب نشان دهنده یک خروجی از آزمون می‌باشد. در این طریق، هر شاخه تلاش می‌کند مشاهدات را در زیر گروه‌هایی به‌صورت دو به دو ناسازگار تقسیم کند. خوبی یک انشعاب بر مبنای انتخاب صفاتی است که بهترین جداسازی نمونه را انجام دهد. نمونه‌ها با موفقیت به دسته‌هایی تقسیم می‌شوند، یعنی تا زمانی که هر یک از دو تا دیگر مسیر تقسیم شدن را دیگر طی ننمایند می‌توان اختلاف معنی‌دار را به‌طور آماری یا زیر گروه‌ها برای ایجاد تقسیم معنی‌دار مشابه بسیار کوچک تولید کرد. چندین الگوریتم تقسیم پیشنهاده شده است. در ردیابی یا بازیابی متقابل خودکار (AID[۷]) بیشترین آماره t معنی دار به‌عنوان هر تحلیل واریانس استفاده می‌شود. در مربع AID chi، آماره chi و طبقه بندی و رگرسیون درختی (CART[۸]) برای استفاده از شاخص تفاوت‌ها بکار رفته است (کوه و لو، ۲۰۰۴). در این مطالعه ما الگوریتم شناخته شده ID3 را بکار می‌بریم. ID3 اندازه گیری مبتنی بر انتروپی است که به‌عنوان اطلاعات سودمند شناخته می‌شود و آن را برای تعیین صفت متمایز کننده انتخابی بکار می‌بریم (هانک و کَمِر، ۲۰۰۰).

تقسیم‌های پی در پی نمونه ممکن است یک درخت بزرگ تولید کند. برخی از شاخه‌های درخت ممکن است بازتاب خلاف قاعده‌ای ازدسته آموزشی باشد مانند ارزش های نادرست یا طرح‌های کلی نگر. به همین دلیل شاخه به میزان مورد نیاز هرس می‌شود. هرس شاخه با حذف گره‌های تقسیم شده در راهی است که در نرخ درستی مدل معنی‌دار نمی‌باشد.

برای تعیین طبقه هدف مشاهده نشده پیشین، وزن صفت هدف دوباره در گره‌های تصمیم‌گیری شاخه آزموده می‌شود. براساس این آزمون یک مسیر پیموده می‌شود که به هدف طبقه در پیشگویی ختم خواهد شد. مزیت اصلی درخت تصمیم گیری این است که یک راه معنی دار با نمایش دانش کسب شده تدارک می‌بینند و قواعد طبقه‌بندی اگر- آنگاه را به آسانی استخراج می‌کند.

۳-۳-۲- شبکه‌های عصبی

شبکه‌های عصبی (NN[۹]) یک فن‌آوری کامل در اجرای یک تئوری و شناخت حوزه‌های کاربردی آن است. یک NN متشکل از تعدادی سلول عصبی و امثال آن به‌صورت واحدهای پردازش به هم پیوسته است. اشتراک در هر اتصال با یک ارزش عددی است که «وزن» نامیده میشود. هر سلول عصبی علامتی از اتصال سلول عصبی می‌گیرد و ترکیب علامت‌های ورودی محاسبه می‌شود. مجموع علامت ورودی برای سلول عصبی j بصورت $u_j = \sum w_{i,j} \times x_i$ است چنانکه x_i علامت ورودی از سلول عصبی i و $w_{i,j}$ وزن اتصال بین سلول عصبی i و سلول عصبی j می‌باشد. اگر قدرت علامت ورودی ترکیب شده بیش از یک آستانه‌ای باشد در این صورت ارزش ورودی به وسیله بخش ویژه انتقال شبکه عصبی انتقال داده می‌شود و در پایان سلول عصبی برانگیخته می‌گردد (هان و کمبر، ۲۰۰۰).

سلول‌های عصبی در لایه‌هایی چیده می‌شوند. هر لایه شبکه متشکل از یک ورودی (اولیه) و یک خروجی (ثانویه) به‌عنوان کوچکترین جزء هر لایه است. در بین ورودی و خروجی لایه ممکن است یک لایه مخفی یا بیشتر وجود داشته باشد. اختلاف انواع NNها در اختلاف تعداد لایه‌های آنها است. طرح‌های سازمان یافته طبیعی (SOM[۱۰]) تنها یک لایه ورودی و خروجی دارند. در حالیکه یک NN پس انتشار یک یا چندین لایه مخفی بیشتر دارد.

بعد از اینکه معماری شبکه مشخص شد، شبکه باید آموزش ببیند. در شبکه‌های پس انتشار یک الگو برای لایه ورودی و محاسبه خروجی نهایی در لایه خروجی بکار می‌رود.خروجی با نتیجه مورد درخواست مقایسه شده و خطاها به عقب NN انتشار می یابد یا وسیله تنظیم وزن اتصال‌ها قرار گیرد. این فرایند تا زمانی تکرار می‌شود که به یک نرخ خطای قابل پذیرش برسیم. شبکه‌های عصبی که هیچ فرضی در مورد صفات مستقل ندارند قابلیت رسیدگی به داده‌های مخدوش یا متناقض را ندارند و پیشنهاد انعطاف‌پذیر مناسب بر مشکلاتی هستند که یک الگوریتم چاره ساز در آن کارایی ندارد. از آنجایی که NNهای پس انتشار بیشترین عمومیت را به‌عنوان یک پیش‌بینی کننده و طبقه‌بندی کننده مشکلات دارد (سهل

Archive of SID
www.SID.ir

و ون کاتاچالام، ۱۹۹۵) ما شبکه‌های عصبی پس انتشار را در این مطالعه گزیدیم.

۳-۳-۳- شبکه‌های گمانه زنی بیزی (شبکه‌های احتمالات شرطی)

طبقه‌بندی بیزی بر مبنای قضیه آماری بیز است. قضیه بیز یک خوشه بندی از احتمالات پس گرا تدارک می‌بیند. بر اساس قضیه بیز، اگر H یک فرضیه باشد- مانند اینکه شئ (یا موضوع) X به طبقه C تعلق داشته باشد- آنگاه احتمال جا افتادن (تحقق) فرضیه به‌صورت $P(H|X) = (P(X|H) * P(H))|P(X)$ می‌باشد.

اگر یک شئ (یا موضوع) X به یک گروهی از طبقات متناوب i تعلق داشته باشد برای تعیین طبقه‌بندی شئ در یک خوشه‌بندی طبقه‌بندی شده بیزی، احتمالات ممکن به‌صورت $P(C_i|X)$ برای تمام طبقات C_i ممکن بوده و بیشترین احتمال تخصیص شئ (یا موضوع) به طبقه‌ای خاص می‌باشد $P(C_i|X)$.

طبقه‌بندی‌های بیزی ساده با فرض حالت مستقل ساخته می‌شود که این در علم آمار از یک ارزش صفت داده شده در یک طبقه که مستقل از ارزش سایر صفات است نتیجه می‌شود. فرض ساده خوشه‌بندی به‌صورت $P(C_i|X)$ است. اگر این فرض درست انگاشته شود طبقه‌بندی ساده بیزی بهترین برآورد صحیح را در مقایسه با بقیه طبقه‌بندی‌ها خواهد داشت. هرچند در بسیاری از موارد این فرض تایید نمی‌شود چون که وابستگی بین صفات می‌تواند وجود داشته باشد.

شبکه‌های گمانه زنی بیزی (BNN[11]) اجازه گزارش مقادیر وابسته زیرمجموعه‌ای از صفات را می‌دهد. یک BNN نمودار ناچرخه‌ای (مارپیچی) مستقیم است که هرگاه یک صفت را نشان می‌دهد و هر فلش نشان‌دهنده وابستگی احتمالی است. اگر یک فلش به‌صورت اگر گره A به گره B رسم شود، آنگاه A پدر B و B زاده A است. در یک شبکه گمانه‌زنی هر متغییر شرطی مستقل از نازاده آن که توسط پدرش داده شده، می‌باشد (هان و کمبر، ۲۰۰۰).

برای هر گره X موجود یک جدول احتمالات شرطی نیز موجود است، هم چنانکه احتمالات شرطی ویژه برای هر ارزش X برای هر ترکیب ارزش‌های ممکن پدرانشان نیز موجود است (توزیع شرط $P(X|Parents(X))$. احتمال یک مجموعه $n,(X_1, X_2, X_3, ..., X_n)$ توزیع دارند که به این شکل است.

ساختار یک شبکه را می‌توان از پیش تعریف نمود یا از داده‌ها استنباط کرد به منظور طبقه‌بندی یک گره از میان گره‌ها می‌توان طبقه گره را تعریف نمود. شبکه توانایی خوشه‌بندی احتمال هر طبقه متناوبی را دارد.

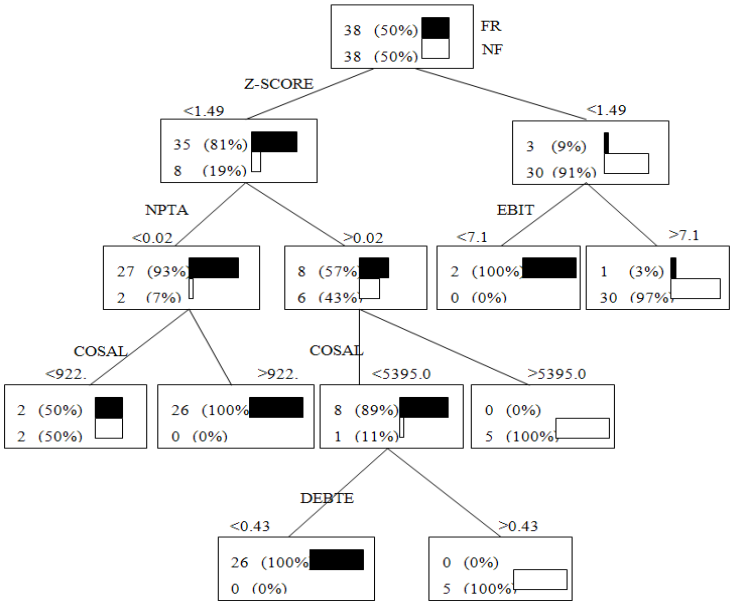
۴- آزمون‌ها و نتایج تحلیل‌ها

سه مدل متنوع که هر کدام بر مبنای روش‌های مختلف بود ساخته شد، مدل درخت تصمیم‌گیری با استفاده از نرم‌افزار ویرایش تحقیقاتی سیپنیا[12] بنا شد. مدل با سطح اطمینان ۹۵٪ ساخته شد. ما تمام نمونه‌ها را به‌عنوان دسته آموزشی استفاده کردیم. شکل شماره (۲) نشان دهنده ساختار ایجاد شده درخت تصمیم‌گیری است.

مدل در برابر داده‌های آموزشی آزمون شد و تا طبقه بندی ۷۳ مورد به‌طور صحیح مدیریت شود (عملکرد کلی برابر ۹۶٪). در این مورد درخت تصمیم گیری تمام موارد فاقد تقلب (۱۰۰٪) و ۳۵ مورد از ۳۸ مورد تقلب را بطور صحیح طبقه بندی کرد.

همانطور که در شکل شماره (۲) مشاهده می‌شود، الگوریتم، متغییر امتیاز Z را به‌عنوان اولین جداساز استفاده نموده است. ۳۵ مورد از ۳۸ مورد شرکتهای متقلب نشان داده شده ارزش امتیاز Z کم قابل توجهی داشتند (۱/۴۹ > امتیاز Z). از آنجائیکه آلتمن ارزش امتیاز Z برابر ۱/۸۱ را نقطه شناسایی پریشانی مالی برای شرکت‌های تولیدی ایالات متحده آمریکا مطرح کرده است (آلتمن، ۲۰۰۱)، ما می‌توانیم استنتاج کنیم شرکت‌هایی با پریشانی مالی در نمونه ما گرایش بیشتری به دستکاری صورت‌های مالی یشان را دارند.

Archive of SID
www.SID.ir



شکل شماره ۲: ساختار ایجاد شده درخت تصمیم‌گیری

در مرحله دوم جداسازی دو متغیر سهیم در سودآوری (NPTA و EBIT) بکار رفته است. شرکتهای بدون تقلب با امتیاز Z بالا نشان دهنده سودآوری بالا در حالیکه شرکت‌های متقلب با امتیاز Z کم نشان دهنده سودآوری پایین می‌باشد. جدول شماره (۳) متغیرهای جداساز نمود یافته در درخت تصمیم‌گیری را شرح می‌دهد.

جدول شماره ۳: تفکیک متغیرها در درخت تصمیم‌گیری

The spilitting variables
Variables
Z-SCORE
NPTA
EBIT
COSAL
DEBTEQ

در آزمون دوم مدل شبکه عصبی را ساختیم. نرم افزار ناکلاس۷[۱۳] با شبکه غیرخطی برای طبقه‌بندی را به منظور ساخت یک شبکه پیش خور مفهومی چند لایه بکار برده شد. بعد از آزمون تعدادی از طرح‌های متنوع، و انجام آموزش‌های مقدماتی، وضعیتی با یک لایه مخفیِ محتوی پنج گره مخفی را برگزیدیم.

شبکه انتخاب شده دوم، با استفاده از تمامی نمونه‌ها آموزش دید و در مقابل دسته آموزشی آزمون شد. شبکه موفق شد تمام موارد را به درستی طبقه‌بندی کند؛ بدین گونه که به موفقیت ۱۰۰ درصدی در عمل دست یافت. متاسفانه نرم افزار نمی‌توانست یک ارتباط شفاف از وزن‌های سیناپس‌های ارتباطی برای ما مهیا کند، بنابراین برآوردی از میزان اهمیت هر متغیر ورودی ممکن نبود.

در آزمایش سوم، یک شبکه گمانه زنی بیزی (BBN) را توسعه دادیم. نرم افزاری که استفاده کردیم BN با قدرت پیش‌بینی بود. این نرم افزار ظرفیت آموزش یک طبقه‌بندی کننده را با داده‌ها دارد. متعلقات الگوریتمی که برپا شده دسته‌ای از الگوریتمهای مبتنی بر آزمون مستقل شرطی است و نیاز به تنظیم گره نیست (جنگ و گرینر ۲۰۰۱). به اقتضای محدودیت‌های نرم افزار ما باید مقادیر را به‌صورت نظری اعمال کنیم. پس از آزمون روش‌های نظری مختلف (با پهنا و ژرفای مساوی) ما روش نظری بر نگری شده را انتخاب کردیم. برخلاف سایر روش‌های نظری، مبنای نظری درگاشت محور[۱۴] در رده‌بندی اطلاعاتی بکار می‌رود. این ساختار شباهتی را که فاصله تعریف شده ممکن است برای کمک به بهبود طبقه‌بندی صحیح ایجاد کند را می‌سازد (هان و کمبر، ۲۰۰۰).

از این رو برای آموزش شبکه گمانه زنی، تمام نمونه‌ها را به‌عنوان دسته آموزشی استفاده و بعد از آموزش شبکه، آن را در مقابل دسته آموزشی آزمون نمودیم. شبکه به درستی ۷۲ مورد را طبقه‌بندی نمود (عملکرد ۹۵٪). در این مورد مدل ۳۷ مورد تقلب (۹۷٪) و ۳۵ مورد فاقد

Archive of SID
www.SID.ir

تقلب (۹۲٪) را به درستی طبقه‌بندی نمود. شکل شماره (۴) ساختار شبکه گمانه زنی را نشان می‌دهد.

همانطور که در شکل شماره (۴) می‌توان دید، شبکه گمانه زنی به منظور وفق دادن تعمیم یافتگی بیشتر، ظاهری با ملاحظه به انگیزه‌های تحریف صورت‌های مالی مدنظر داشته است. براساس شبکه تقلبات گزارش داده شده به شدت وابستگی به ورودی متغیرهای امتیاز Z- ،DEBTEQ ،NPTA، SALTA و WCTA دارند. هرکدام ازاین متغیرها ادعای اختلاف یک جنبه از وضعیت مالی شرکت را دارد. امتیاز Z- بازتابی از پریشانی مالی است؛ DEBTEQ به اهرم اشاره دارد؛ NPTA نشان از قابلیت سودآوری؛ SALTA به عملکرد فروش اشاره دارد و WCTA نشان از نقدشوندگی است.

براین اساس شبکه گمانه زنی برای ثبت وابستگی بین صورتهای مالی تحریف شده، تعدادی از جنبه‌های مالی شایان توجه هر شرکت را مدنظر قرار داده است. جدول شماره (۵) متغیرهای انتخاب شده برای شبکه گمانه‌زنی را نشان می‌دهد.

جدول شماره ۵: متغیرهای منتخب

Variables
Z-SCORE
NPTA
DEBITQ
SALTA
WCTA

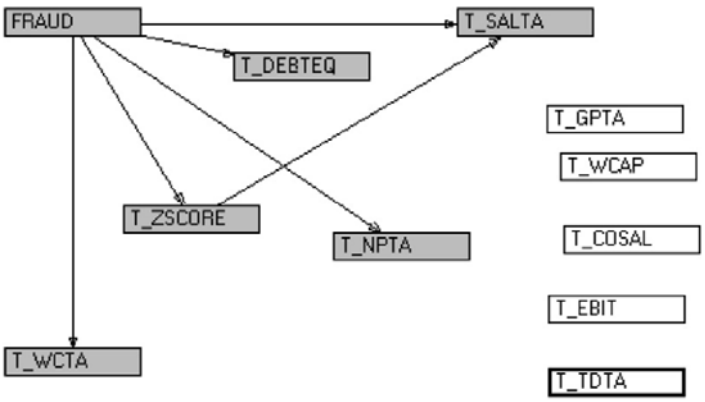
جدول شماره (۶) عملکرد سه مدل را بر نمونه آموزشی تشریح می‌کند. نتایج اشاره به این دارد که مدل NN به‌صورت کاملاً کارا قدرت تشخیص بین شرکتهای دارای FFS و بدون FFS را دارد و بعد از آن مدلهای BBN و ID3 قرار دارند.

جدول شماره ۶: عملکرد سه مدل منتخب داده کاوی جهت کشف تقلب

Performance against the training set			
Model	Fraud (%)	Non-fraud (%)	Total (%)
ID3	92.1	100.0	96.2
NN	100.0	100.0	100.0
BBN	97.4	92.1	94.7

۴-۱- (روایی) اعتبار مدل‌ها

استفاده از دسته آموزشی به منظور تعیین برآوردی از توان اجرایی یک مدل به‌عنوان یک مبنا معرفی شده است. در بسیاری از موارد در مدل‌ها گرایش به خلاصه‌سازی نمونه در برابر آموزش وجود دارد، (داده‌هایی که قرار است برارزش شوند). برای برطرف نمودن چنین مواردی یک زیربنای اجرایی از مدل قبل از اینکه الگوها دیده شوند برآورد می‌شود. چندین رویکرد اعتباربخشی به مدل مانند تقسیم نمونه‌ها به آموزشی و بخش کنارگذاشته نمونه، ۱۰ لایه متقاطع اعتباربخشی و کنار گذاشتن یکی برای اعتبار بخشی وجود دارد.



شکل شماره ۴: ساختار شبکه گمانه زنی

Archive of SID
www.SID.ir

اگرچه سه بسته نرم‌افزاری که در تحقیق مورد استفاده گردید متضمن (دربردارنده) قابلیت‌های اعتباربخشی بودند، دنبال کردن فرآیند یک اعتباربخشی عمومی در معنای روش‌شناسی (متدولوژی) و داده برای هر سه بسته نرم‌افزاری غیرممکن بود. در نتیجه باید به شکل دستی نمونه‌ها را تقسیم و نمونه آموزشی و نمونه اعتبارسنجی ساخته می‌شد.

در این راستا، از رویکرد اعتباربخشی ۱۰- لایه متقاطع طبقه‌بندی شده انتخاب شد که نمونه‌ها را در ۱۰ لایه تقسیم می کرد. در رویکرد طبقه‌بندی شده هر لایه از تعداد مساوی موارد تقلب و فاقد تقلب تشکیل شده است. برای هر لایه، مدل با استفاده ۹ لایه باقیمانده آموزش می‌دید و با لایه کنار گذاشته شده آزمون می‌شد. در پایان میانگین عملکرد محاسبه گردید. در جدول شماره (۷) اعتباربخشی ۱۰ لایه متقاطع عملکرد سه مدل خلاصه شده است.

جدول شماره ۷: نتایج حاصل از اعتبار بخشی مدل‌های منتخب

Performance against the training set			
Model	Fraud (%)	Non-fraud (%)	Total (%)
ID3	75.0	72.5	73.6
NN	82.5	77.5	80.0
BBN	91.7	88.9	90.3

همانطور که انتظار می‌رفت، نسبت درستی دسته اعتباربخشی کمتر از نسبت درستی دسته آموزشی بود. هرچند عملکرد هر کدام از سه مدل به شکل شایان توجهی متفاوت بود. مدل درخت تصمیم‌گیری که در طبقه‌بندی خود به درستی ۹۶٪ نمونه‌های آموزشی را مدیریت نموده بود، کاهش شایان توجهی در درصد صحت طبقه‌بندی وقتی در مقابل نمونه اعتباربخشی قرار گرفت، از خود نشان داد. مدل به درستی ۷۳/۶٪ کل نمونه‌ها را طبقه‌بندی نمود یعنی ۷۵٪ موارد تقلب و ۷۲/۵٪ موارد فاقد تقلب. مدل شبکه عصبی که از یک عملکرد خالص ۱۰۰٪ با دسته آموزشی بهره‌مند بود، به درستی ۸۰٪ طبقه‌بندی نمونه اعتباربخشی را مدیریت نمود یعنی ۸۲/۵٪ از موارد تقلب و ۷۷/۵٪ موارد فاقد تلقب. در پایان شبکه گمانه‌زنی بیزی که کمترین صحت را در دسته آموزشی داشت، با موفقیت ۹۱/۷٪ موارد آموزشی و ۸۸/۹٪ موارد فاقد تقلب و در کل ۹۰/۳٪ دسته اعتباربخشی را به درستی طبقه‌بندی نمود.

در یک ارزیابی تطبیقی از عملکرد مدل‌ها می‌توانیم نتیجه بگیریم که شبکه گمانه زنی بیزی کاربردی‌ترین نسبت به دو مدل دیگر است و به‌عنوان: برجسته‌ترین طبقه‌بندی کننده صحیح رسیده است. شبکه‌های عصبی به‌عنوان: عملکرد بالا به‌طور رضایت‌بخش رسیده و در پایان عملکرد درخت تصمیم‌گیری به‌عنوان: ترجیح کمتر مطرح شد.

در ارزیابی عملکرد یک مدل سایر ملاحظات مهم، نرخ‌های خطاهای نوع ۱ و ۲ است. خطای نوع ۱ چنین پذیرفته شده: وقتی یک شرکت متقلب به‌عنوان فاقد تقلب طبقه‌بندی شود. خطای نوع ۲ چنین پذیرفته شده است: وقتی یک شرکت فاقد تقلب به‌عنوان متقلب طبقه‌بندی شود. خطاهای نوع ۱ و ۲ هزینه‌های متفاوتی دارند. طبقه‌بندی یک شرکت متقلب به‌عنوان یک شرکت فاقد تقلب ممکن است ما را به سمت تصمیمات نادرست هدایت کند که عواقب وخیم اقتصادی در بر خواهد داشت. طبقه بندی اشتباه یک شرکت فاقد تقلب ممکن است به علت صرفه هزینه زمانی بررسی بیشتر شود. هر چند هیچ مدلی به کاهش نرخ‌های هر دو نوع خطا نائل نشده است. یک مدل فرض رجحان خطای نوع ۱ نسبت به خطای نوع ۲ آن را دارد. در آزمون، تمام مدل‌ها نرخ خطای نوع ۱ کمتری را به نمایش گذاشتند. شبکه عصبی مصنوعی بیشترین اختلاف بین نرخ خطای نوع ۱ و ۲ را ارائه نمود.

۵- نتیجه‌گیری و بحث

داده کاوی به‌عنوان یکی از ده دانش در حال توسعه، پل ارتباطی میان علم آمار، علم کامپیوتر، هوش مصنوعی، الگوشناسی، فراگیری ماشین و بازنمایی بصری داده می‌باشد که کاربرد گسترده‌ایی در زمینه‌های پزشکی، صنعت، مهاجرت، هتلداری، مدیریت ریسک و بالاخص کشف تقلب دارد، بطوریکه امروزه فعالیت حسابرسی باید چیرگی بر موارد رو به فزونی تقلب مدیران

Archive of SID
www.SID.ir

باشد، نیازمند ابزارهای مؤثر می‌باشد. تکنیک‌های داده کاوی که ادعا می‌شود ظرفیت‌های پیشرفته‌ای در طبقه‌بندی و پیش‌بینی دارند، توان تسهیل کنندگی برای حسابرسان در انجام وظیفه کشف تقلب مدیریتی را دارند. ادعای این مطالعه بررسی سودمندی و مقایسه عملکرد سه ترفند (تکنیک) داده کاوی در کشف صورت‌های مالی متقلبانه به وسیله استفاده از داده‌های مالی عمومی بوده است. مدل‌های بکار رفته درخت تصمیم‌گیری، شبکه‌های عصبی، شبکه‌های گمانه زنی بیزی بود.

نتایج حاصل از آزمون‌ها که مؤید نتایج تحقیقات پیشین بود به این اشاره داشت که داده‌های صورت‌های مالی، محتوی اطلاعاتی جهت کشف تحریف است. از این گذشته به‌طور نسبی فهرست نسبت‌های مالی کوچک به شدت تاثیر تعیین‌کننده‌ای در طبقه‌بندی دارند. این آگاهی، زوج همزمان[15] (نقطه اتصال) الگوریتم‌های داده کاوی است که می‌تواند ظرفیت مدل در رسیدن به صحت و توجه به نتایج طبقه‌بندی را تدارک ببیند.

مطالعه ارائه شده مبنی بر تجربیات تحقیقات حسابرسی و حسابداری را که توان ایجاد بهترین تمایز موارد FFS را دارد، مشارکت داده است. این مطالعه همچنین متغیرهای معینی از اطلاعات عمومی آماده که حسابرسان باید زمان حسابرسی بیشتری به آن اختصاص دهند را توصیه کرده است. چهارچوب روش شناسی پیشنهادی مورد استفاده می‌تواند کمکی برای حسابرسان داخلی و خارجی و مالیاتی و سایر اعتباردهندگان گزارشات، اشخاص حقیقی و حقوقی، سرمایه‌گذاران، معامله گران سهام، شرکت‌های وکالت، تحلیل گران اقتصادی، موسسات رتبه‌دهی اعتباری و سامانه (سیستم) بانکی باشد. برای حسابرسی حرفه‌ای این مطالعه می‌تواند در مساعدت برای شناسایی (روش‌های) پاسخگویی در برابر کشف FFS سودمند باشد.

از نظر کارایی مدل شبکه گمانه‌زنی بیزی با بهترین عملکرد، به میزان ۹۰/۳٪ صحت طبقه بندی از نمونه اعتباری در فرایند اعتباربخشی ۱۰- لایه متقاطع دست یافت. نسبت صحت مدل شبکه بندی عصبی و مدل درخت تصمیم‌گیری به ترتیب ۸۰٪ و ۷۳/۶٪ بود. نسبت خطای نوع ۱ برای هر سه مدل کم بود.

شبکه گمانه زنی بیزی وابستگی بین تحریف و نسبت‌های بدهی به ارزش ویژه، سود خالص به مجموع دارایی‌ها، فروش به مجموع دارایی‌ها، سرمایه در گردش به مجموع دارایی‌ها و امتیاز Z را آشکار نمود. هر کدام از این نسبت‌ها بازگشتی از اختلاف مورد انتظار از وضعیت مالی یک شرکت به ترتیب همچون اهرم قابلیت سودآوری، عملکرد فروش، نقد شوندگی و پریشانی مالی است. مدل درخت تصمیم‌گیری در ابتدا تحریف را با پریشانی مالی همبسته کرد، از این رو امتیاز Z را به‌عنوان اولین مرحله جداسازی استفاده نمود.

همانگونه که معمولاً اتفاق می افتد، این مطالعه می‌تواند به‌عنوان سنگ زیر بنای تحقیقات آتی استفاده شود. یکی از اختلافات مهم بین آزمون‌ها این بود که در مدل BBN به علت محدودیت‌های نرم افزار داده‌ها به‌صورت نظری بکار گرفته شد. داده‌های نظری عناصر تاثیرگذار در طرح کلی دخیل در هزینه، فقدان بعضی از اطلاعات است. پیشبرد تحقیق نیاز به تعیین عناوین داده‌های نظری، بر عملکرد مدل و عناوین الگوریتم‌های نظری تاثیر زیادی دارد. همچنین تحقیق نیازمند آزمون چگونگی این موضوع است که کدام مدل‌های داده کاوی (DM) عملکرد بهتری نسبت به بقیه ترفندها (تکنیک‌ها) دارند.

ورودی‌های خط دهنده مبنایی در این تحقیق، مبتنی بر تحقیق Krikos, spathis & manolopulos بوده است که تنها متشکل از نسبت‌های مالی بودند. غنی‌سازی ورودی‌های خط دهنده با اطلاعات کیفی مانند صلاحیت حسابرسان پیشین با ترکیب هیات مدیره می‌تواند درستی نسبت را افزایش دهد. از این گذشته مطالعه ویژه یک صنعت می‌تواند نتایج ویژه آن را آشکار نماید. امید است مطالعه ارائه شده در این مقاله موجبات تحقیقات بیشتر در مورد عناوین مهم این موضوع باشد.

فهرست منابع

۱) حسینی نسب، مرضیه و همکاران (۱۳۹۰)، «کشف تقلب در نسخ دارویی به کمک روش‌های داده کاوی و ترکیب اطلاعات Fwltext»

www.SID.ir
Archive of SID

۲) رهنمای رودپشتی، فریدون، همتی، هدی (۱۳۸۹)، «روش‌شناسی پژوهش‌های دانشگاهی تقلب و سوء استفاده‌های مالی» مجموعه مقالات اولین کنفرانس پیشگیری از تقلب و سوء استفاده‌های مالی، تهران، سالن برج میلاد،آذر ماه.

۳) یاس (۱۳۹۰)، «کانون هدایت فکر: داده کاوی چیست؟»

4) http://www.civilica.com/paper-IDMCO2-IDMCO2-035.html

5) 4. Chan, P.K & et.al [1999] "Distributed data mining in credit card fraud detection" IEEE intelligent systems, November / December, pp.67-74

6) 5. Jans, M. Lybaert, N. & Vanhoof, K [2007] "Data mining for fraud detection: Toward an Improvement on internal control systems" http://www.fraud

7) 6. Kirkos, E, Spathis, C. and Manolopoulos, Y [2007] "Data mining techniques for the detection of fraudulent financial statement" expert systems with applications, vol.32, pp.995-1003.

8) 7. Lee, W. and Stolfo, S.J. [1998] "Data mining approaches for intrusion detection" sysmposium San Antonio, Texas, January 26-29.

* از آقای محمود انصار دانشجوی کارشناسی ارشد رشته مدیریت مالی واحد علوم و تحقیقات دانشگاه آزاد اسلامی که در ترجمه بخش‌هایی از مقاله مرجع همکاری داشته‌اند، تشکر و قدردانی می‌شود

یادداشت‌ها

[1]. Data Mining
[2]. Fraudulent Financial Statement (FFS)
[3]. Artificial Intelligence
[4]. Association of Certified Fraud Examiners
[5]. American Institute of Certified Public Accountants
[6]. Computer-Assisted Audit tools & Techniques (CAATTS)
[7]. Automatic Interaction Detection
[8]. Classification and Regression Trees
[9]. Neural Networks
[10]. Self-organizing maps
[11]. Bayesian Belief Networks
[12]. Sipina Research Edition
[13]. Nuclass7
[14]. Entropy-based
[15]. couple

Archive of SID

www.SID.ir