

کاربست روش‌های داده کاوی به منظور ارتقای عملکرد

تشخیص فرار مالیاتی

محسن دستگیر^۱

مریم غریبی^۲

تاریخ دریافت: ۱۳۹۴/۵/۱۹ تاریخ پذیرش: ۱۳۹۴/۱۱/۱۴

چکیده

این پژوهش به بررسی کاربست روش‌های داده کاوی به منظور ارتقای عملکرد تشخیص فرار مالیاتی می‌پردازد. داده کاوی، فرایند کشف اطلاعات نامعلوم، ناشناخته و پنهان از یک پایگاه داده است و روشی است منحصر به فرد برای یافتن حقایق جدید و روابط بین داده‌های موجود که به وسیله صاحب نظران کشف نشده است. در این پژوهش سودمندی داده کاوی مبتنی بر قواعد وابستگی به عنوان ابزاری برای تشخیص فرار مالیاتی به کار گرفته شده است. جامعه آماری این پژوهش کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌باشد که بر اساس شرایط در نظر گرفته شده برای انتخاب نمونه به روش حذف سامانمند، ۱۲۵ شرکت در دوره زمانی ۱۳۸۳ تا ۱۳۹۰ انتخاب گردید. در این پژوهش از ۲۸ متغیر مالی و غیرمالی در قالب ۹ طبقه به منظور ایجاد مدل استفاده شد. قواعد وابستگی با به کارگیری الگوریتم پیشینار برای تشخیص فرار مالیاتی شرکت‌ها استفاده شد. بدین منظور داده‌ها به طور تصادفی به سه دسته آموزش، اعتبارسنجی و آزمون تقسیم شدند. نتایج پژوهش نشان داد که روش‌های داده کاوی مبتنی بر قواعد وابستگی با ایجاد دو مدل با درصد صحت ۹۱٪ بر روی داده‌های آموزش، با درصد صحت ۸۸٪ بر روی داده‌های اعتبارسنجی و با درصد صحت ۸۶٪ بر روی داده‌های آزمون توانسته است موفق به تشخیص فرار مالیاتی گردد.

واژه‌های کلیدی: فرار مالیاتی، قواعد وابستگی، داده کاوی

۱. دکترای حسابداری و عضو هیأت علمی حسابداری دانشگاه آزاد اسلامی واحد علوم و تحقیقات اصفهان

۲. دانشجوی کارشناسی ارشد حسابداری دانشگاه آزاد اسلامی واحد مبارکه (نویسنده مسئول) gharibimaryam563@yahoo.com

۱- مقدمه

فعالیت‌های فرار مالیاتی، همواره از مسائل مهم و مسئله ساز در هر کشوری بوده است. به منظور تحقق بخشیدن به منافع حاصل از هزینه‌های ارزشمند، هرچند محدود، برای تشخیص فرار مالیاتی، مقامات مالیاتی نیاز به مصرف منطقی منابع خود را دارند (وو و همکاران، ۲۰۱۲). هوش تجاری^۱ به طور کلی و داده کاوی^۲ به طور خاص، ابزارهای مؤثری برای افزایش کارایی و اثربخشی تشخیص فرار از پرداخت مالیات هستند (فدایرو و همکاران، ۲۰۰۸). ایالات متحده آمریکا یکی از اولین کشورها در اعمال روش‌های داده کاوی برای تشخیص فرار مالیاتی و در نتیجه جبران مالیات پرداخت نشده بوده است (هوور، ۲۰۰۹). سانجینی (۲۰۰۴) گزارش می‌دهد که ایالت تگزاس با استفاده از یک سیستم هوش تجاری، ۳۶۲ میلیون دلار زیان مالیاتی دولت را بازیافت نموده است، به همین دلیل مقامات مالیاتی در تگزاس به شدت متعهد به استفاده از روش‌های داده کاوی برای کشف گزارش مالیاتی مشکوک هستند.

در این پژوهش، داده کاوی مبتنی بر قواعد وابستگی^۳ دنبال می‌شود. قواعد وابستگی یکی از کاربردی‌ترین و شناخته شده ترین روش در زمینه داده کاوی است. این قواعد، لزوم وقوع برخی از متغیرها را در صورت وقوع برخی دیگر از متغیرها، تضمین می‌کنند. هدف از این روش، جستجو و یافتن بیشترین ارتباط معنی دار در بین مقادیر متغیرها یا ویژگی‌هاست (داتا و توماس، ۱۹۹۹) که در این پژوهش با پیاده‌سازی الگوریتم پیشینار^۴ برای کشف الگوها و روابط بین متغیرهای مربوط به عوامل تأثیرگذار بر فرار مالیاتی به کار گرفته شده است (آگراوال و سریکانت، ۱۹۹۴).

یکی از موارد کاربردی فرار مالیاتی که با توجه به مفاد ماده ۱۵۸ قانون مالیات‌های مستقیم، در این پژوهش مورد استفاده قرار می‌گیرد این است که طبق مفاد این ماده قانونی، سازمان امور مالیاتی می‌تواند اظهارنامه‌های مؤدیان مالیاتی را بدون رسیدگی قبول کند و فقط تعدادی از آنها را به صورت نمونه گیری رسیدگی کند. مؤدیانی که اظهارنامه آنها در اجرای این ماده مورد رسیدگی قرار می‌گیرد، در صورتی که سود مشمول مالیات قطعی با سود مشمول مالیات ابرازی آنها بیش از ۱۵٪ اختلاف داشته باشد، طبق ماده ۱۹۴ همان قانون مشمول جریمه خواهند بود. مفاد ماده ۱۵۸ ق.م.م تاکنون برای رسیدگی به اظهارنامه‌های مالیاتی تسلیمی به وسیله اشخاص حقیقی استفاده شده است. با اجرایی شدن قانون مالیات بر ارزش افزوده از اول مهر ماه سال ۱۳۸۷ و مروری بر برنامه‌های راهبردی سازمان امور مالیاتی کشور، این امر

1. Business Intelligence (BI)
2. Data Mining
3. Association Rules
4. Apriori

آشکار می‌گردد که فرایند تشخیص مالیات به سمت خوداظهاری مؤدیان مالیاتی در حرکت است و به نظر می‌رسد در آینده‌ای نه چندان دور، مفاد ماده ۱۵۸ ق.م.م در رسیدگی به اظهارنامه‌های مالیاتی تسلیمی به وسیله اشخاص حقوقی نیز تسری یابد. به هر روی، در این پژوهش میزان ۱۵٪ موضوع ماده ۱۹۴ ق.م.م به اشخاص حقوقی نیز تسری داده شده تا مبنایی برای قضاوت در مورد وجود یا نبود فرار مالیاتی برای شرکت‌های مورد آزمون فراهم آورد. با توجه به مطالب مطرح شده این پژوهش به دنبال پاسخ به این پرسش است که آیا به کارگیری قواعد وابستگی برای کشف الگوها و قواعد مرتبط بین متغیرهای تأثیرگذار بر فرار مالیاتی شرکت‌ها امکانپذیر است؟

این مقاله در سه بخش ساماندهی شده است: در بخش اول مبانی نظری و پیشینه پژوهش مطرح شده است در بخش دوم روش پژوهش شامل جامعه و نمونه آماری؛ فرضیه‌ها؛ مدل پژوهش و روش پژوهش بیان شده است، پس از آن در بخش سوم به ارائه یافته‌ها؛ نتیجه‌گیری و پیشنهادات پرداخته شده است.

۲- مبانی نظری پژوهش

از آنجا که فناوری داده کاوی از قابلیت‌های پیش بینی و طبقه بندی فراوانی برخوردار است می‌تواند فرایند تصمیم‌گیری در مسائل مالی را تسهیل نماید. کاربرد روش‌های داده کاوی با توجه به مطالعات مرتبط و ماهیت آنها، می‌تواند طیف گسترده‌ای شامل پیش بینی ورشکستگی، تخمین ریسک اعتباری، وضعیت تداوم فعالیت، درماندگی مالی، پیش بینی عملکرد واحد تجاری و انواع تقلب را در بر گیرد، بنابراین در این پژوهش جهت تشخیص فرار مالیاتی به کار گرفته شده است.

متغیرهای مستقل مورد استفاده در این پژوهش، نسبت‌های مالی و غیرمالی هستند که از صورت‌های مالی شرکت‌های نمونه استخراج شده‌اند که در این پژوهش با توجه به مطالعات انجام شده قبلی و به دلایل زیر انتخاب گردیدند.

طبق نظر دجو و همکاران (۱۹۹۶) یکی از انگیزه‌های مهم تقلب در گزارشگری مالی حداقل کردن بدهی مالیاتی است. همچنین البوت و ویلینگهام (۱۹۸۰) عنوان می‌کنند که تقلب، همان دستکاری هوشیارانه صورت‌های مالی است که به منظور کاهش مالیات بر سود انجام می‌شود. بنابراین با توجه به اینکه انگیزه شرکت‌ها از فرار مالیاتی حداقل کردن بدهی مالیاتی و کاهش سود مشمول مالیات خود است می‌توان نتیجه گرفت که فرار مالیاتی شرکت‌ها نیز از تقلب و دستکاری صورت‌های مالی ناشی می‌شود. بدین منظور، برخی از پژوهش‌های انجام شده در خصوص استفاده از نسبت‌های مالی و غیرمالی جهت کشف تقلب مورد بررسی قرار خواهد گرفت.

اسپاتیس (۲۰۰۲) با استفاده از رگرسیون لجستیک^۱ مرحله ای، نشان داد که نسبت‌های مالی از توانایی بالقوه در کشف گزارشگری مالی متقلبانانه برخوردارند. کامینسکی و همکاران (۲۰۰۴) با استفاده از رگرسیون حداقل مربعات، به این نتیجه دست یافتند که نسبت‌های مالی توان توضیح دهندگی زیادی به منظور کشف تقلب در گزارشگری مالی دارند.

کوتسیاننیتس و همکاران (۲۰۰۶) با استفاده از شیوه‌های مبتنی بر یادگیری ماشین، بیان کردند که نسبت‌های مالی شرکت‌ها از توانایی بالقوه جهت کشف تقلب در گزارشگری مالی برخوردارند. گراو و باسیلیکو (۲۰۰۸) با استفاده از الگوی آماری پروبیت، نشان دادند که عوامل مالی و عوامل راهبری شرکتی در شناسایی تقلب در گزارشگری مالی مؤثرند. احمد و همکاران (۲۰۰۹) با استفاده از رگرسیون حداقل مربعات و نظریه هزینه‌های سیاسی بیان کردند که اندازه شرکت و کیفیت حسابرسی دارای رابطه معنادار منفی با تقلب در گزارشگری مالی است. سامرز و سویی (۱۹۹۸) دریافتند که حسابرسان می‌توانند از طریق توجه بیشتر به مبادلات داخلی، احتمال موفقیت را در کشف تقلب در گزارشگری مالی افزایش دهند.

آنساه و همکاران (۲۰۰۲)، با استفاده از رگرسیون حداقل مربعات اندازه مؤسسه حسابرسی، وضعیت تصدی حسابرس و تجربه کاری حسابرسان را پیش بینی کننده‌های خوبی از احتمال کشف تقلب بیان کردند. ارنست و یانگ (۲۰۰۳) دریافتند که بیش از نصف افرادی که مرتکب تقلب در گزارشگری مالی می‌شوند از گروه مدیریت شرکت هستند. کمیته سازمان‌های حامی کمیسیون تردوی^۲ در گزارشی تحت عنوان تقلب در گزارشگری مالی در سال ۱۹۹۹ عنوان کرد که ۷۲ درصد از موارد مورد بررسی به وسیله مدیر عامل و ۴۲ درصد به وسیله مدیر مالی صورت گرفته است. بیشتر موارد مطالعه شده حاکی است که مدیر عامل، اعضای هیئت مدیره و مدیران مالی نقش بیشتری در ایجاد تقلب داشته‌اند.

بیزلی و همکاران (۲۰۰۰) به بررسی تفاوت‌های راهبری شرکتی در نمونه‌ای از شرکت‌های متقلب و غیرمتقلب پرداختند. آنها دریافتند که شرکت‌های متقلب با احتمال کمتری کمیته حسابرسی دارند؛ شرکت‌های غیرمتقلب در مقایسه با شرکت‌های متقلب از مدیران بیرونی بیشتری در کمیته حسابرسی بهره می‌برند و واحد حسابرسی داخلی در شرکت‌های متقلب با احتمال کمتری وجود دارد.

۳- پیشینه پژوهش

در ادامه برخی از مهم‌ترین پژوهش‌های صورت گرفته در خصوص موضوع پژوهش بیان می‌شود.

1. Logistic Regression
2. Committee of Sponsoring Organizations of the Treadway Commission

پژوهش‌های خارجی

گاماچیس و همکاران (۲۰۱۲) در پژوهشی با عنوان «مدل پشتیبان تصمیم‌گیری برای وصول درآمدهای مالیاتی در یونان» مدل پارامتریک مارکوف^۱ را برای فرار از پرداخت مالیات به وسیله شرکت‌های یونانی پیشنهاد کردند. آنها این مدل را برای شناسایی پارامترهای مالیاتی مؤثر و منصفانه به کار گرفتند و رفتار مطلوب شرکت‌ها را با توجه به این پارامترها محاسبه کردند. آنها بیان نمودند که با توجه به اینکه مالیات بر مجموع درآمد در یونان یک اولویت بالا در نظر گرفته شده است، به عنوان گزینه مالیاتی جایگزین، فرار از پرداخت مالیات را افزایش می‌دهد.

راویسانکار و همکاران (۲۰۱۱) در پژوهشی با عنوان «تشخیص تقلب در صورت‌های مالی و متغیرهای انتخاب شده با استفاده از روش‌های داده کاوی» به شناسایی شرکت‌هایی که به تقلب در صورت‌های مالی متوسل می‌شوند پرداختند. آنها روش‌های داده کاوی را بر روی مجموعه‌ای از داده‌های شرکت‌ها در دو مرحله آزمون کردند. در مرحله اول با انتخاب ۱۸ ویژگی، شبکه‌های عصبی احتمالاتی^۲ از بین تمامی روش‌ها برتر بود و در مرحله دوم نیز با انتخاب ۱۰ ویژگی، شبکه‌های عصبی احتمالاتی و برنامه نویسی ژنتیک^۳ با حاشیه دقت برابر نسبت به سایر روش‌ها برتر بودند.

هاشم زاده و همکاران (۲۰۱۰) در پژوهشی با عنوان «تقلب مالیاتی شرکت‌ها و حسابرسی مطلوب» یکی از موارد تقلب را ادعای بیش از حد برای تخفیف مالیات بر ارزش افزوده کالاهای صادراتی می‌دانند و نشان دادند که وجود تقلب نمی‌تواند تصمیمات خروجی واقعی شرکت و سیاست‌های مالیاتی دولت را تحت تأثیر قرار دهد. بنابراین باید منابع حسابرسی را برای شناسایی شرکت‌های درگیر در تقلب و نیز برای واسطه‌هایی که مدارک جعلی عرضه می‌کنند مورد استفاده قرار داد. با این وجود، آنها با مفروضات معقول نشان دادند که این منابع باید برای شناسایی شرکت‌ها و نه واسطه‌ها متمرکز شود.

چن و همکاران (۲۰۰۹) در پژوهشی با عنوان «استفاده از روش‌های داده کاوی و شبکه‌های عصبی مصنوعی^۴ برای پیش‌بینی مدل اضطراب مالی» از شبکه‌های عصبی مصنوعی برای پیش‌بینی تقلب استفاده کردند و نشان دادند که شبکه‌های عصبی مصنوعی از دقت بالایی در کشف تقلب برخوردارند. کایرکاس و همکاران (۲۰۰۷) در پژوهشی با عنوان «روش‌های داده کاوی برای تشخیص صورت‌های مالی جعلی» به بررسی اثربخشی روش‌های طبقه‌بندی داده کاوی در ۷۶ شرکت تولیدی یونانی برای

1. Parametric Markov model
2. Probabilistic Neural Network
3. Genetic Programming
4. Artificial Neural Networks

شناسایی شرکت‌هایی که صورت‌های مالی جعلی منتشر می‌کنند پرداختند. آنها بیان کردند که در تشخیص تقلب مدیریت، حساب‌رسان می‌توانند با استفاده از روش‌های داده‌کاوی در کارشان تسهیل ایجاد کنند. آنها در پژوهش خود سودمندی درخت‌های تصمیم‌گیری^۱، شبکه‌های عصبی و شبکه‌های باور بیزین^۲ را در شناسایی صورت‌های مالی جعلی بررسی کردند.

پژوهش‌های داخلی

فلاحتی و همکاران (۱۳۹۱) در پژوهشی با عنوان «مدل سازی اقتصاد سایه‌ای و تخمین فرار مالیاتی در ایران با استفاده از شبکه عصبی مصنوعی» با استفاده از تکنیک شبکه عصبی مصنوعی به بررسی و مدل سازی حجم اقتصاد پنهان و محاسبه میزان فرار مالیاتی در اقتصاد ایران پرداختند. نتایج برآورد مدل شبکه عصبی نشان می‌دهد که متغیرهای حجم دولت، نرخ خود اشتغالی و بازبودن اقتصادی از مهم‌ترین عوامل مؤثر بر حجم اقتصاد پنهان در ایران می‌باشد.

عباسی و همکاران (۱۳۹۱) در پژوهشی با عنوان «لزوم توسعه به‌کارگیری روش‌های داده‌کاوی در حسابداری و حسابرسی» بیان کردند که امروزه پژوهشگران با توجه به توان بالای فناوری داده‌کاوی و هوش مصنوعی در پردازش پایگاه‌های داده‌های بزرگ و یافتن الگوهای پیچیده و غیرخطی در آنها، از روش‌های آماری به سمت روش‌های هوش مصنوعی پیش می‌روند. نتایج آنها نشان می‌دهد که استفاده از روش‌ها و روش‌های داده‌کاوی حتی به حساب‌رسان این امکان را می‌دهد تا با استفاده از حداقل زمان و هزینه، گزارشی را ارائه کنند که مستند به روش‌های علمی بوده و از اتکاپذیری و اطمینان بالایی برخوردار باشد. بنابراین، لزوم توسعه و به‌کارگیری روش‌های داده‌کاوی در حوزه علوم مالی را، ضرورتی اجتناب‌ناپذیر دانستند.

باقرپور و ولاشانی و همکاران (۱۳۹۱) در پژوهشی با عنوان «پیش‌بینی گزارش حسابرس مستقل در ایران: رویکرد داده‌کاوی» با استفاده از سه روش داده‌کاوی شامل درخت تصمیم، شبکه‌های عصبی مصنوعی و رگرسیون لجستیک به پیش‌بینی گزارش حسابرس مستقل پرداختند. نتایج پژوهش حاکی از آن است که میانگین دقت مدل حاصل از تکنیک درخت تصمیم از دو روش دیگر بیشتر بوده و این روش توانسته است گزارش حسابرسی را با میانگین ۸۸٫۶۴٪ پیش‌بینی کند.

موسوی و همکاران (۱۳۸۸) در پژوهشی با عنوان «فرار مالیاتی در نظام مالیات بر ارزش افزوده: یک مدل نظری» به بررسی عوامل اقتصادی مؤثر بر شکل‌گیری و گسترش پدیده فرار مالیاتی در نظام مالیات

1. Decision Trees
2. Bayesian Belief Networks

بر ارزش افزوده پرداختند. نتایج نشان داد که با افزایش پرداخت‌های هزینه‌ای مشمول مالیات بر ارزش افزوده، فرار مالیاتی افزایش می‌یابد و به عبارتی تمکین مالیاتی کاهش خواهد یافت. همچنین، افزایش در احتمال رسیدگی و کشف فرار مالیاتی و افزایش نرخ جریمه، برای مواردی که پرونده مالیاتی رسیدگی نمی‌شود، افزایش تمکین مالیاتی و در نتیجه کاهش فرار مالیاتی را موجب خواهد شد.

فراست و همکاران (۱۳۸۶) در پژوهشی با عنوان «داده کاوی و کاربرد آن در تصمیم‌گیری» به تشریح تکنولوژی داده کاوی در حوزه هوش محاسباتی و فرآیند استاندارد میان صنعتی داده کاوی (CRISP-DM) پرداختند و یک طبقه بندی از کاربردهای این تکنولوژی در حوزه تصمیم‌گیری‌های کسب و کار ارائه دادند. آنها بیان می‌کنند که توان و قدرت تکنولوژی داده کاوی و مدل‌های متنوعی که این فناوری از آنها استفاده می‌نماید، ابزار مناسبی جهت اخذ تصمیمات مختلف در زمینه‌هایی مانند پیش‌بینی قیمت سهام، انتخاب سبد سهام و کشف تخلفات است.

۴- جامعه و نمونه آماری

جامعه آماری این پژوهش شامل کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران در دوره زمانی ۱۳۸۳ تا ۱۳۹۰ می‌باشد. داده‌های نمونه از شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران جمع آوری شده است. داده‌های مربوط به واحد تجاری مشکوک به عنوان موارد فرار از پرداخت مالیات شناسایی شد. روش نمونه برداری در این پژوهش روش حذف سامانمند است؛ لذا نمونه انتخابی با در نظر گرفتن شرایط زیر انتخاب شد.

۱. اطلاعات پرونده شرکت‌ها به صورت کامل و پیوسته در دوره بررسی، در دسترس باشد.
۲. به لحاظ افزایش قابلیت مقایسه، دوره مالی آنها منتهی به پایان اسفند ماه هر سال باشد.
۳. سود قطعی آنها در ساها ی ۱۳۸۳ لغایت ۱۳۹۰ مشخص باشد.

۵- فرضیه پژوهش

فرضیه پژوهش به این صورت بیان می‌شود: تشخیص فرار مالیاتی شرکت‌ها با استفاده از روش‌های داده کاوی مبتنی بر قواعد وابستگی امکانپذیر است.

۵-۱- تبیین و اندازه‌گیری متغیرهای پژوهش

متغیرهای این پژوهش شامل متغیرهای وابسته و مستقل است که در ادامه به توصیف آنها پرداخته خواهد شد.

تبیین و اندازه‌گیری متغیر وابسته

متغیر وابسته در این پژوهش، عبارت است از: وجود یا عدم وجود فرار مالیاتی در یک واحد تجاری که بر طبق رابطه شماره ۱ محاسبه می‌شود:

$$\% \Delta = (TAXIN - ACCIN / ACCIN) \times 100 \quad (1)$$

که در آن:

$\% \Delta$: درصد اختلاف سود مشمول مالیات ابرازی و سود مشمول مالیات قطعی شرکت i در سال t است.

ACCIN: سود مشمول مالیات ابرازی شرکت i در پایان سال مالی (t) .

TAXIN: سود مشمول مالیات قطعی شرکت i در پایان سال مالی (t) .

بنابراین طبق مفاد ماده ۱۹۴ ق.م.م در صورتی که سود مشمول مالیات قطعی شده با سود مشمول مالیات ابرازی واحد تجاری بیش از ۱۵٪ اختلاف داشته باشد، به عنوان موارد فرار مالیاتی واحد تجاری در نظر گرفته خواهد شد.

تبیین و اندازه‌گیری متغیرهای مستقل

متغیرهای مستقل این پژوهش در قالب ۹ طبقه جهت دستیابی به هدف پژوهش انتخاب شد که عبارتند از:

۱. ساختار سرمایه: متغیرهایی که در این طبقه قرار می‌گیرند، شامل نسبت بدهی، بدهی‌های بلندمدت به دارایی‌ها، بدهی‌های بلندمدت به حقوق صاحبان سهام و نسبت پوشش بهره می‌باشد.
۲. سودآوری: برای محاسبه سودآوری از نسبت سود عملیاتی به فروش خالص، نسبت حاشیه سود، سود عملیاتی به دارایی‌ها، بازده کل دارایی‌ها و بازده حقوق صاحبان سهام استفاده شده است.
۳. عملکرد: جهت محاسبه عملکرد شرکت از نرخ رشد فروش، نسبت دارایی‌ها به فروش خالص و عملکرد شرکت (در صورتی که عملکرد شرکت سود باشد عدد یک و در صورت زیان عدد صفر اختصاص خواهد یافت) استفاده شده است.

رشد فروش: نسبت رشد فروش را نشان می‌دهد و از فرمول زیر برای محاسبه آن استفاده شده است.

$$\frac{\text{دوره قبل - فروش دوره جاری}}{\text{فروش دوره قبل}} \quad (2)$$

۴. نقدینگی: نسبت‌هایی که در این گروه قرار دارند، توانایی شرکت در انجام تعهدات مالی جاری را اندازه‌گیری می‌کنند، که شامل نسبت وجه نقد به بدهی‌های جاری، نسبت آنی، نسبت جاری و نسبت

- سرمایه در گردش به دارایی‌ها هستند (نوروش و همکاران، ۱۳۸۶).
۵. جریان نقد عملیاتی: از تقسیم میزان وجه نقد خالص فعالیت‌های عملیاتی دوره جاری بر ارزش دفتری کل دارایی‌ها به دست می‌آید.
۶. نسبت فروش به حقوق صاحبان سهام: مقدار فروش تقسیم بر حقوق صاحبان سهام واحد تجاری.
۷. راهبری شرکتی: در یک نگاه کلی راهبری شرکتی، شامل ترتیبات حقوقی، فرهنگی و نهادی است که سمت و سوی حرکت و عملکرد شرکت‌ها را تعیین می‌نماید (نمازی و کرمانی، ۱۳۸۷).
- ۷-۱. تغییر حسابرس: اگر حسابرس نسبت به سال گذشته تغییر یافته عدد یک، در غیر این صورت مقدار صفر اختصاص یافت.
- ۷-۲. نوع حسابرس: اگر حسابرسی به وسیله سازمان حسابرسی صورت پذیرد عدد یک، در غیر این صورت عدد صفر اختصاص یافت.
- ۷-۳. دوگانه بودن مسؤلیت مدیر عامل: به حالتی اطلاق می‌شود که سمت مدیر عامل و رئیس یا نایب رئیس هیأت مدیره در فرد واحدی جمع باشد که در این حالت عدد صفر، در غیر این صورت مقدار یک اختصاص یافت.
- ۷-۴. تغییر مدیر عامل: اگر مدیر عامل نسبت به سال گذشته تغییر یافته باشد عدد صفر، در غیر این صورت عدد یک اختصاص یافت.
- ۷-۵. استقلال هیأت مدیره: از نسبت مدیران غیرموظف به کل اعضای هیأت مدیره شرکت محاسبه شده است.
- ۷-۶. سهامداران نهادی: به اشخاص حقوقی اطلاق می‌شود که تعداد قابل ملاحظه‌ای از سهام عادی شرکت را در اختیار داشته باشند. مؤسسات سرمایه‌گذاری، نهادهای عمومی غیردولتی و یا سایر شرکت‌های تجاری از جمله سهامداران نهادی محسوب می‌شوند. این متغیر عبارت است از درصد تجمعی سهام شرکت که به وسیله نهادهای عمده سرمایه‌گذاری نگه‌داری می‌شود به کل سهام منتشره و در دست سهامداران شرکت.
- ۷-۷. نوع گزارش حسابرس: چنانچه اظهار نظر حسابرس غیراستاندارد (اظهار نظر مشروط، مردود یا عدم اظهار نظر) باشد مقدار آن صفر، در غیر این صورت مقدار یک اختصاص یافت.
۸. اندازه شرکت: از طریق لگاریتم طبیعی میانگین دارایی‌ها محاسبه شده است.
۹. تاریخ درج در بورس: این متغیر نشان‌دهنده فاصله زمانی بین تاریخ درج در بورس اوراق بهادار تهران تا

سال مورد رسیدگی است.

۱۰. ماهیت مالکیت: اگر مالکیت اولین سهام‌دار عمده (بیش از ۵۰٪ سهام شرکت) متعلق به دولت باشد عدد یک، در غیر این صورت مقدار صفر اختصاص یافت.

جدول (۱) - متغیرهای پژوهش

متغیرهای مستقل			
ردیف	طبقه متغیر	نام متغیر	اختصار
۱	ساختار سرمایه	نسبت بدهی	DA
۲		نسبت بدهی‌های بلندمدت به دارایی‌ها	LDA
۳		نسبت بدهی‌های بلندمدت به حقوق صاحبان سهام	LDE
۴		نسبت پوشش بهره	EBITI
۵	سودآوری	نسبت سود عملیاتی به فروش خالص	OINS
۶		نسبت حاشیه سود	NIR
۷		نسبت سود عملیاتی به دارایی‌ها	EBITA
۸		نسبت بازده کل دارایی‌ها	NIA
۹		نسبت بازده حقوق صاحبان سهام	NIE
۱۰	عملکرد	نرخ رشد فروش	GRG
۱۱		نسبت دارایی‌ها به فروش خالص	ANS
۱۲		سود یا زیان	GOR
۱۳	نقدینگی	نسبت وجه نقد به بدهی‌های جاری	CCD
۱۴		نسبت آنی	AACD
۱۵		نسبت جاری	CACD
۱۶		نسبت سرمایه در گردش به دارایی‌ها	WCA

متغیرهای مستقل				
ردیف	طبقه متغیر	نام متغیر	اختصار	
۱۷	جریان نقد عملیاتی	تقسیم وجه نقد خالص فعالیت‌های	OCFTA	
		عملیاتی بر ارزش دفتری		
		کل دارایی‌ها		
۱۸	فروش خالص به حقوق صاحبان سهام	مقدار فروش تقسیم بر حقوق صاحبان سهام	SE	
۱۹	راهبری شرکتی	تغییر حسابرس	ATC	
۲۰		نوع حسابرس	ATT	
۲۱		دوگانه بودن مسؤولیت مدیر عامل	DRD	
۲۲		تغییر مدیر عامل	DCh	
۲۳		استقلال هیأت مدیره	IB	
۲۴		سهامداران نهادی	ISh	
۲۵		نوع گزارش حسابرس	POP	
۲۶		اندازه شرکت	لگاریتم طبیعی میانگین دارایی‌ها	LNA
۲۷		سایر عوامل	تاریخ درج در بورس	AGE
۲۸			ماهیت مالکیت	INDU

۵-۲- طرح آزمون فرضیه‌ها

در این پژوهش جهت کشف الگوها و روابط بین متغیرهای تأثیرگذار بر فرار مالیاتی، قواعد وابستگی با به‌کارگیری الگوریتم پیشینار مورد بررسی قرار گرفت که به صورت یک تابع غیرخطی پارامتری بر روی مجموعه‌ای از داده‌ها به کار برده شد. قواعد وابستگی به شرح زیر است:

قواعد وابستگی

یکی از مهم‌ترین بخش‌های داده کاوی، کشف قواعد وابستگی در پایگاه داده است. از جمله کارهایی که کاوش وابستگی برای ما انجام می‌دهد می‌توان به پیدا کردن وابستگی‌ها و همبستگی‌های موجود در بین داده‌ها، یافتن الگوهایی که غالباً در بین داده‌ها وجود دارند و همچنین پیدا کردن یک سری ساختار سببی در بین اقلام و اشیای موجود در پایگاه داده‌های تعاملی و رابطه‌ای اشاره کرد (وو و همکاران، ۲۰۱۲).

در یادگیری قواعد وابستگی، پشتیبان (پوشش)^۱ و اطمینان (دقت)^۲ دو پارامتر مهم برای قضاوت در مورد اعتبار قواعد وابستگی هستند. این دو پارامتر به شرح زیر تعریف شده اند:

تعریف ۱. پشتیبان: احتمال وجود همزمان A_1 و A_2 به صورت همزمان در تراکنش است.

تعریف ۲. اطمینان: میزان وابستگی یک قلم کالای خاص را به دیگری بیان می کند.

$$\text{support}(A_1, A_2, \dots, A_j, A_k) \equiv P(A_1, A_2, \dots, A_j, A_k) \quad (۳)$$

$$= \frac{\#(\text{K - dependent Attributes})}{\#(\text{total Reports})}$$

$$\text{Confidence}(A_1, A_2, \dots, A_j \rightarrow A_k) \equiv P(A_k | A_1, A_2, \dots, A_j) \quad (۴)$$

برای معنی داری قواعد وابستگی، پشتیبان و اطمینان باید بزرگتر از آستانه قابل قبول باشد (کائو و همکاران، ۲۰۰۷). در این راستا الگوریتم پیشینار به کار برده شد.

الگوریتم پیشینار

هدف در این الگوریتم، پیدا کردن بزرگترین مجموعه اقلام است که حداقل پشتیبان و اطمینان را رعایت کند. دو فرض زیر در این الگوریتم در نظر گرفته می شود (آگراوال و سریکانت، ۱۹۹۴):

۱. هر زیر مجموعه از یک مجموعه قلم تکرار شونده، تکرار شونده است (یعنی اگر فرضاً مجموعه $\{b, c, a\}$ تکرار شونده باشد، آنگاه مجموعه $\{b, a\}$ نیز تکرار شونده است).

۲. هر فوق مجموعه از یک مجموعه قلم تکرار نشونده، تکرار نشونده است (یعنی اگر فرضاً مجموعه $\{b, a\}$ تکرار شونده نباشد، آنگاه مجموعه $\{c, b, a\}$ نیز تکرار شونده نیست).

۶- روش پژوهش

این پژوهش، به شناسایی عوامل مؤثر بر فرار مالیاتی پرداخته است و سودمندی داده کاوی مبتنی بر قواعد وابستگی به عنوان ابزاری برای کشف فرار مالیاتی مورد بررسی قرار گرفت. لذا از لحاظ هدف، بنیادی تجربی و از جنبه ماهیت، همبستگی است.

۷- یافته های پژوهش

نتایج حاصل از اجرای الگوریتم پیشینار بر روی سه مجموعه داده، شامل داده های آموزش^۳، اعتبارسنجی^۴

1. Support
2. Confidence
3. Training
4. Validation

و آزمون^۱ با استفاده از نرم افزار سی شارپ مورد بررسی قرار گرفته است.

۷-۱- نتایج فاز مدل‌سازی با استفاده از داده‌های آموزش

داده‌های آموزش برای استخراج الگوهای فرار مالیاتی با استفاده از الگوریتم پیشینار به کار گرفته شد. اولین کاوش بر روی مجموعه داده‌ای انجام گرفت که در مرحله آماده سازی اولیه شکل گرفته بود. با اعمال روش قواعد وابستگی نتایجی به دست آمد که جهت نیل به هدف پژوهش، نسبت به بررسی داده‌ها و انتخاب مجدد خصیصه‌ها اقدام گردید. پس از انجام بررسی‌های لازم و ارزیابی نتایج، متغیر سود یا زیان (GOR) که یکی از متغیرهای مستقل پژوهش است به دلیل تولید این متغیر به صورت متغیر خروجی به تعداد زیاد و درصد صحت پایین برای ایجاد مدل‌های فرار مالیاتی، تصمیم به طبقه‌بندی این متغیر بین رده صفر و یک نموده و فرآیند داده کاوی دوباره تکرار شد.

انتخاب پشتیبان و اطمینان

پشتیبان و اطمینان برای ایجاد قواعد وابستگی به ترتیب در طیف وسیع ۱۰-۵۰٪ و ۸۰-۹۰٪ جهت داده‌های آموزش در نظر گرفته شد. تعداد قواعد به دست آمده از داده کاوی بر روی این مجموعه داده در جدول ۲ نشان داده شده است:

جدول (۲) - نتایج داده کاوی بر روی داده‌های آموزش

اطمینان			پشتیبان
۹۰٪	۸۵٪	۸۰٪	
۹۶۲۶	۱۲۳۵۰	۱۴۰۸۶	۱۰٪
۲۱۱	۲۴۸	۲۸۸	۲۰٪
۱۷	۲۰	۲۹	۳۰٪
۳	۳	۵	۴۰٪
۰	۰	۰	۵۰٪

همانطور که در جدول ۲ مشاهده می‌شود با پشتیبان ۵۰٪ قانونی تولید نمی‌شود، به همین دلیل پشتیبان در محدوده‌ای کمتر از ۵۰٪ در نظر گرفته شد. همچنین برای ایجاد قواعد مطلوب و بالا بودن درصد وابستگی بین متغیرها مقدار اطمینان از ۸۰٪ به بالا انتخاب شد. پس از آن، لازم است قواعد بدست آمده ارزیابی

1. Test

گردد تا درصد صحت آنها مشخص شود. بدین منظور از نرم افزار سی شارپ استفاده گردید. با ارزیابی قواعد به دست آمده بر روی داده‌های آموزش، با پشتیبان ۲۰٪ و اطمینان ۸۵٪ برای مجموعه داده‌های آموزش شش قاعده به دست آمد که به شرح جدول ۳ می‌باشند.

جدول (۳) - استخراج قواعد وابستگی بر روی مجموعه داده‌های آموزش

متغیرهای مستقل	متغیر وابسته	پشتیبان %	اطمینان %
LDA=0.1 OINS=0.1	taxevasion=yes <==	۲۳	۱۰۰
LDA=0.1 OINS=0.1 GOR=0.1	taxevasion=yes <==	۲۲	۱۰۰
OINS=0.1 GOR=0.1	taxevasion=yes <==	۳۳	۹۲
EBITA=0.1 OCFTA=0.1	taxevasion=yes <==	۴۵	۹۲
EBITA=0.1 OCFTA=0.1 POP=0	taxevasion=yes <==	۳۵	۹۱
NIR=0.1 OCFTA=0.1	taxevasion=yes <==	۲۴	۸۶

قواعد وابستگی به صورت اگر و آنگاه تولید می‌شوند به عنوان مثال قاعده اول بدین معنی است که اگر متغیر نسبت بدهی-های بلندمدت به دارایی‌ها (LDA) به میزان ۰/۱، نسبت سود عملیاتی به فروش خالص (OINS) به میزان ۰/۱ باشد آنگاه فرار مالیاتی رخ خواهد داد.

اجرای الگوریتم پیشینار

در این الگوریتم از دو فرض مطرح شده در قبل استفاده شده است.

مرحله ادغام

ابتدا مجموعه قواعد به دست آمده به ترتیب مرتب شد. سپس مجموعه قواعدی که با یکدیگر قابل ادغام و دارای عنصر مشترک بودند جهت رسیدن به نتیجه مطلوب ترکیب شدند. این مجموعه‌ها عبارتند از:

$$L1=LDA=0.1 \text{ OINS}=0.1$$

$$L2=LDA=0.1 \text{ OINS}=0.1 \text{ GOR}=0.1$$

$$L3=OINS=0.1 \text{ GOR}=0.1$$

به این ترتیب، پس از ترکیب قواعد بالا خواهیم داشت:

$$C1=\{LDA=0.1, \text{ OINS}=0.1, \text{ GOR}=0.1\}$$

دومین مجموعه قابل ادغام، قواعد زیر می‌باشند.

$$L4=EBITA=0.1 \text{ OCFTA}=0.1$$

$$L5=EBITA=0.1 \text{ OCFTA}=0.1 \text{ POP}=0$$

$$L6=NIR=0.1 \text{ OCFTA}=0.1$$

به این ترتیب، مجموعه ترکیب شده حاصل به شرح زیر خواهد بود.

$$C2=\{EBITA=0.1, \text{ OCFTA}=0.1, EBITA=0.1, \text{ POP}=0, \text{ NIR}=0.1\}$$

مرحله هرس

C_1 و C_2 مجموعه‌ای از قواعد هستند که هر عضو آن یا مکرر است یا خیر، اما تمام عناصر مکرر در آن قرار دارند. حال تمام عناصر این مجموعه باید بررسی شوند تا مکرر بودن آنها مشخص شود.

$$C1= \{LDA=0.1, OINS=0.1, GOR=0.1\}$$

$$C2= \{EBITA=0.1, \text{ OCFTA}=0.1, EBITA=0.1, \text{ POP}=0, \text{ NIR}=0.1\}$$

بنابراین پس از ترکیب و ادغام مجموعه قواعد C_1 و C_2 ، قواعد به شرح زیر جهت ارزیابی آماده شدند.

$$\{LDA=0.1 \text{ OINS}=0.1\}$$

$$\{LDA=0.1 \text{ OINS}=0.1 \text{ GOR}=0.1\}$$

$$\{\text{OINS}=0.1 \text{ GOR}=0.1\}$$

عناصر مجموعه C_1

$$\{LDA=0.1 \text{ GOR}=0.1\}$$

$$\{LDA=0.1\}, \{\text{OINS}=0.1\}, \{\text{GOR}=0.1\}$$

$$\{EBITA=0.1 \text{ OCFTA}=0.1\}$$

$$\{EBITA=0.1 \text{ OCFTA}=0.1 \text{ POP}=0\}$$

$$\{EBITA=0.1 \text{ OCFTA}=0.1\}$$

$$\{EBITA=0.1 \text{ NIR}=0.1 \text{ OCFTA}=0.1\}$$

عناصر مجموعه C_2

$$\{\text{POP}=0 \text{ NIR}=0.1 \text{ OCFTA}=0.1\}$$

$$\{EBITA=0.1\}, \{\text{OCFTA}=0.1\}, \{EBITA=0.1\}, \{\text{POP}=0\}, \{\text{NIR}=0.1\}$$

پس از طی نمودن مراحل ادغام و هرس کردن قواعد، دو قاعده به شرح جدول ۴ به دست آمد که نتیجه نهایی پژوهش می‌باشند. قاعده اول نسبت بدهی‌های بلندمدت به دارایی‌ها (LDA) به میزان ۰/۱، نسبت سود عملیاتی به فروش خالص (OINS) به میزان ۰/۱ و سود یا زیان (GOR) به میزان ۰/۱ می‌باشد و قاعده دوم نسبت سود و زیان قبل از بهره و مالیات به دارایی‌ها (EBITA) به میزان ۰/۱، نسبت خالص

جریان نقد فعالیت‌های عملیاتی بر ارزش دفتری کل دارایی‌ها (OCFTA) به میزان ۰/۱، می‌باشد، که اولین قانون با درصد اطمینان ۱۰۰٪ و پشتیبان ۲۲٪ و دومین قانون با درصد اطمینان ۹۲٪ و پشتیبان ۴۵٪ به عنوان مدلی برای تشخیص فرار مالیاتی انتخاب گردیدند.

جدول (۴) - استخراج قواعد با استفاده از داده‌های آموزش پس از ترکیب و ادغام قواعد با صحت ۹۰٪

اطمینان	پشتیبان	متغیر وابسته	متغیرهای مستقل
۱۰۰	۲۲	taxevasion=yes <==	LDA=0.1 OINS=0.1 GOR=0.1
۹۲	۴۵	taxevasion=yes <==	EBITA=0.1 OCFTA=0.1

به منظور افزایش کارایی قواعد به‌دست آمده با استفاده از داده‌های آموزش، این قواعد همزمان بر روی داده‌های اعتبارسنجی نیز مورد ارزیابی قرار گرفت که در ادامه بیان خواهد شد.

۷-۲- اجرای فاز ارزیابی با استفاده از داده‌های اعتبارسنجی

اعتبارسنجی به طور متوسط، میزان صحت پیش بینی را اندازه‌گیری می‌کند. بعد از محاسبه متوسط صحت پیش بینی قواعد، قواعد با صحت پیش بینی پایین، هرس شدند. دسته بندی قواعد باقیمانده برای نشان دادن الگوی فرار از پرداخت مالیات، مورد استفاده قرار گرفت. به عبارتی در این مرحله فرایند ارزیابی بر روی داده‌های اعتبارسنجی انجام شد و نسبت به حذف و اضافه نمودن قواعد اقدام گردید که نتایج بدست آمده از ارزیابی این مجموعه داده در جدول ۵ نشان داده شده است.

جدول (۵) - ارزیابی قواعد به‌دست آمده بر روی مجموعه داده‌های اعتبارسنجی با درصد صحت ۸۸٪

نتایج ارزیابی	نتایج ارزیابی	نتایج ارزیابی %
قاعده ۱	۲۱۶	۸۶
قاعده ۲	۱۵۰	۶۰
قاعده ۱ و ۲	۲۲۱	۸۸

قواعد به‌دست آمده با استفاده از داده‌های آموزش به‌طور همزمان بر روی داده‌های اعتبارسنجی نیز ارزیابی شدند و همچنین ترکیب قواعد و حذف و اضافه نمودن قواعد در این مرحله نیز انجام گرفت که طبق جدول ۵ نتایج ارزیابی از قاعده اول به‌طور جداگانه با درصد صحت ۸۶٪ و قاعده دوم با درصد صحت ۶۰٪ بدست آمد که با اضافه نمودن قاعده دوم درصد صحت به ۸۸٪ رسید که نرخ نهایی به‌دست آمده پس از ارزیابی

قواعد به وسیله داده‌های اعتبارسنجی می‌باشد.

۷-۳- اجرای فاز پیاده‌سازی با استفاده از داده‌های آزمون

در این مرحله قواعد استخراج شده با استفاده از داده‌های آموزش و ارزیابی شده با استفاده از داده‌های اعتبارسنجی مورد آزمایش قرار گرفت تا قابل استفاده بودن نتایج مشخص گردد. بنابراین برای محاسبه نرخ صحت جهت داده‌های آزمون، قواعد وابستگی به دست آمده از مراحل قبل به کار گرفته شد که نتایج به شرح جدول ۶ می‌باشد.

جدول (۶) - ارزیابی قواعد به دست آمده بر روی مجموعه داده‌های آزمون با درصد صحت ۸۶٪

نتایج ارزیابی نسبت به کل	نتایج ارزیابی نسبت به کل	
۸۶٪	۲۱۵	وجود فرار مالیاتی
۱۴٪	۳۴	عدم وجود فرار مالیاتی
۱۰۰٪	۲۵۰	جمع

همانطور که در جدول ۶ مشاهده می‌شود نرخ صحت به دست آمده ۸۶٪ می‌باشد که نشان دهنده مطلوب بودن قواعد تولید شده و نرخ صحت به دست آمده می‌باشد.

۸- نتیجه گیری و پیشنهادات

نتایج حاصل از انجام این پژوهش بیانگر این است که امکان تشخیص فرار مالیاتی شرکت‌ها با استفاده از روش‌های داده کاوی مبتنی بر قواعد وابستگی امکان‌پذیر است و مدل به دست آمده با درصد صحت ۹۱٪ بر روی داده‌های آموزش، ۸۸٪ بر روی داده‌های ارزیابی و ۸۶٪ بر روی داده‌های آزمون می‌تواند منجر به تشخیص فرار مالیاتی گردد. طبق مدل حاصله با اهمیت‌ترین متغیرهای مستقل جهت تشخیص فرار مالیاتی در مدل اول، نسبت بدهی‌های بلندمدت به دارایی‌ها (LDA) از طبقه ساختار سرمایه، نسبت سود عملیاتی به فروش خالص (OINS) از طبقه سود آوری، سود یا زیان (GOR) از طبقه عملکرد به میزان ۰/۱ و در مدل دوم نسبت سود و زیان قبل از بهره و مالیات به دارایی‌ها (EBITA) از طبقه سود آوری، نسبت خالص جریان نقد فعالیت‌های عملیاتی بر ارزش دفتری کل دارایی‌ها (OCFTA) از طبقه جریان نقد عملیاتی به میزان ۰/۱، می‌باشد. درصد صحت این مدل در جدول ۷ ارائه شده است.

جدول (۷) - نتایج حاصل از اجرای قواعد وابستگی با نرخ صحت ۸۹٪ و خطای ۱۱٪

روش	متغیر وابسته طبقه‌بندی	داده‌های آموزش		داده‌های اعتبارسنجی		داده‌های آزمون		جمع
		درصد	تعداد	درصد	تعداد	درصد	تعداد	
تعداد وابستگی	صحیح	۹۱٪	۴۵۵	۸۸٪	۲۳۱	۸۹٪	۸۹۱	
	اشتباه	۰۹٪	۴۵	۱۲٪	۲۹	۱۱٪	۱۰۹	
	جمع	۱۰۰٪	۵۰۰	۱۰۰٪	۲۵۰	۱۰۰٪	۱۰۰۰	

طبق نتایج به دست آمده میانگین صحت مدل حاصل از این تکنیک ۸۹٪ جهت طبقه‌بندی صحیح فرار مالیاتی می‌باشد. درصد صحت داده‌های آموزش جهت ایجاد مدل فرار مالیاتی ۹۱٪، درصد صحت داده‌های اعتبارسنجی ۸۸٪ می‌باشد که با دقت بالاتری فرار مالیاتی را نسبت به داده‌های آزمون (۸۶٪) طبقه‌بندی می‌کند. نتایج حاصل از این پژوهش وجود رابطه بین نسبت‌های مالی شرکت‌ها و فرار مالیاتی با استفاده از روش‌های داده کاوی را نشان می‌دهد که متغیرهای انتخاب شده جهت انجام پژوهش مطابق با یافته‌های اسپاتیس (۲۰۰۲)، کامینسکی و همکاران (۲۰۰۴)، کوتسیانتیس و همکاران (۲۰۰۶)، کایرکاس و همکاران (۲۰۰۷)، گراو و باسیلیکو (۲۰۰۸)، چن (۲۰۰۹)، احمد و همکاران (۲۰۰۹)، رابسانکار و همکاران (۲۰۱۱)، صفرزاده (۱۳۸۹) و باقرپور و لاشانی (۱۳۹۱) در به کارگیری روش‌های داده کاوی با

استفاده از نسبت‌های مالی جهت تشخیص، پیش بینی تقلب و صورت‌های مالی جعلی می‌باشد. همچنین یافته‌های این پژوهش در خصوص به-کارگیری قواعد وابستگی مطابق با یافته‌های وو و همکاران (۲۰۱۲) جهت تشخیص فرار مالیات بر ارزش افزوده می‌باشد. بنابراین مطابق با پژوهش‌های صورت گرفته قواعد وابستگی از روش‌های داده کاوی توانسته است در تشخیص فرار مالیاتی به کار گرفته شود. در خصوص محدودیت‌های پژوهش می‌توان گفت:

۱. به علت استفاده از روش حذف سامانمند در انتخاب نمونه آماری، واحدهای تجاری انتخابی، حجم محدودی از شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران را تشکیل می‌دهند؛ بنابراین در تعمیم نتایج به کلیه واحدهای تجاری که در حال حاضر در بورس اوراق بهادار تهران پذیرفته شده‌اند باید با احتیاط عمل شود.

۲. تعیین مقدار واقعی فرار مالیاتی انجام شده امکان‌پذیر نیست، لذا در پژوهش حاضر به منظور محاسبه فرار مالیاتی از تعمیم مفاد ماده ۱۵۸ ق. م. م و ماده ۱۹۴ ق. م. م استفاده شده است. در صورتی که از سایر مدل‌ها به منظور محاسبه فرار مالیاتی استفاده شود، ممکن است نتایج متفاوتی حاصل شود.

۳. مالیات عملکرد ۱۳۹۱ شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران تا تاریخ انجام این پژوهش، قطعی نگردیده است، لذا دوره پژوهش حاضر محدود به عملکرد ۱۳۸۳ الی ۱۳۹۰ شده است.

با توجه به نتایج به دست آمده از این پژوهش، می‌توان پیشنهادهایی به شرح زیر ارائه داد: اولین و مهمترین پیشنهاد به سازمان امور مالیاتی است: توصیه می‌شود نتایج این پژوهش را در جهت رسیدگی ممیزان در امور تشخیص مالیات مدنظر قرار داده و ابزارهای نوین را جهت بهبود روند تشخیص در کنار روش‌های دستی و سنتی استفاده نمایند. دومین پیشنهاد به حساب‌رسان صورت‌های مالی است: به آنها توصیه می‌شود نتایج این پژوهش را جهت رسیدگی صورت‌های مالی شرکت‌ها در کنار تجربه حرفه‌ای خود مدنظر قرار دهند. در ادامه موضوعات زیر جهت انجام پژوهش‌های جدید به پژوهشگران پیشنهاد می‌گردد.

۱. از سایر تکنیک‌های داده کاوی به منظور تشخیص فرار مالیاتی استفاده گردد.

۲. از سایر مدل‌های محاسبه فرار مالیاتی برای بررسی میزان فرار مالیاتی استفاده گردد (در این پژوهش از تعمیم ماده ۱۵۸ ق. م. م و مفاد ماده ۱۹۳ ق. م. م استفاده شده است).

فهرست منابع

۱. باقرپور ولاشانی، محمدعلی، محمدجواد ساعدی، علی مشکانی و مصطفی باقری (۱۳۹۱). پیش بینی گزارش حسابرس مستقل در ایران: رویکرد داده کاوی. دهمین همایش حسابداری ایران، دانشگاه الزهراء.
۲. عباسی، ابراهیم، یوسف قنبری، معصومه نجاری و سعید شیپوریان (۱۳۹۱). لزوم توسعه به کارگیری روش های داده کاوی در حسابداری و حسابرسی. دهمین همایش حسابداری ایران، دانشگاه الزهراء.
۳. فراست، علیرضا، سیدمحمدرضا ناصرزاده، محمدتقی تقوی فرد (۱۳۸۶). داده کاوی و کاربرد آن در تصمیم گیری. نشریه دانش مدیریت ۲۰ (۷۹): ۳ - ۱۴.
۴. فلاحتی، علی، مینو نظیفی و سحر عباسپور (۱۳۹۱). مدل سازی اقتصاد سایه ای و تخمین فرار مالیاتی در ایران با استفاده از شبکه عصبی مصنوعی. فصلنامه تحقیقات توسعه اقتصادی (۶): ۳۳-۵۸.
۵. موسوی جهرمی، یگانه، فرهاد طهماسبی بلداجی و نرگس خاکی (۱۳۸۸). فرار مالیاتی در نظام مالیات بر ارزش افزوده: یک مدل نظری. فصلنامه تخصصی مالیات ۱۷ (۵): ۲۷ - ۳۸.
۶. نمازی، محمد و احسان کرمانی (۱۳۸۷). تاثیر ساختار مالکیت بر عملکرد شرکت های پذیرفته شده در بورس اوراق بهادار تهران. بررسی های حسابداری و حسابرسی (۵۳): ۸۳ - ۱۰۰.
7. Agrawal, R., R. Srikant (1998). Fast Algorithm for Mining Association Rules, In Proc. of the VLDB Conference, Santiago, Chile, September 1994. Expanded Version Available As Ibm Research Report, RJ9839, June 1994.
8. Ahmed, N., J. Mohd-Nor & N. Mohd-Saleh. (2009). Fraudulent Financial Reporting and Company's Characteristics. Accounting Perspectives 7 (3): 189-226.
9. Ansah, S. O., G. D. Moyes, P. B. Oyelere & D. Hay. (2002). An Empirical Analysis of the Likelihood of Detecting Fraud in New Zealand. Managerial Auditing Journal, 17 (4): 192-204.
10. Beasley, M. S., J. V. Carcello, D. R. Hermanson & P. D. Lapidés. (2000). Fraudulent Financial Reporting: Consideration of Industry Traits and Corporate Governance Mechanisms. Accounting Horizons 14 (4): 441- 454.
11. Chen, Wei-Sen, & Yin-Kuan Du. (2009). Using Neural Networks and Data Min-

- ing Techniques for the Financial Distress Prediction Model, *Expert Systems with Applications*, 36, 4075–4086.
12. Datta, A., & Thomas, H. (1999). The Cube Data Model: A Conceptual Model and Algebra for On-Line Analytical Processing in Data Warehouses. *Decision Support Systems*, 27, 289-301.
13. Dechow, P. A., R. G. Sloan & A. P. Sweeney (1996). Causes and Consequences of Earnings Manipulation: an Analysis of Firms subject to Enforcement Actions by the SEC. *Contemporary Accounting Research* 13 (1): 1-36.
14. Elliot, R. & J. Willingham. (1980). *Management Fraud: Detection and Deterrence*. Petrocelli, New York. NY.
15. Ernest & Yang (2003). *Fraud, the Unmanaged Risk: English Global Survey*. London.
16. Fadairo, S. A., Williams, R., Trotman, R., & Onyekelu-Eze, A. (2008). Using Data Mining to Ensure Payment Integrity. *Journal of Government Financial Management*, 57, 22–24.
17. Grove, H. & E. Basilio. (2008). Fraudulent Financial Reporting Detection: Key Ratios Plus Corporate Governance Factors, *Int. Studies of Mgt. & Org.*, 38 (3): 10-42.
18. Goumagias, N.D., D. Hristu-Varsakelis, & A. Saraidaris (2012). A Decision Support Model for Tax Revenue Collection in Greece. *Decision Support Systems*, 53 (1): 76-96.
19. Gupta, R. (2008). Tax Evasion and Financial Repression. *Journal of Economics and Business*, 60, 517–535.
20. Hashimzade, N., Z. Huang, & G. Myles. (2010). Tax Fraud by Firms and Optimal Auditing. *International Review of Law and Economics*, 30 (1): 10-17.
21. Hoover, J. N. (2009). States use BI, Data Warehousing to Recoup Unpaid Taxes. *Intelligent Enterprise*, 12.

22. Kaminsky, K. A., T. S. Wetzel & L. Guan (2004). Can Financial Ratios Detect Fraudulent Financial Reporting? *Managerial Auditing Journal* 19 (1): 15-28.
23. Kirkos, E., Ch. Spathis, & Y. Manolopoulos (2007). Data Mining Techniques for the Detection of Fraudulent Financial Statements, *Expert Systems with Applications*, 32 995-1003.
24. Kotsiantis, S., E. Koumanakos, D. Tzelepis & V. Tampakas. (2006). Forecasting Fraudulent Financial Statements using Data Mining. *Proceedings of World Academy of Science, Engineering and Technology* 12: 284-289.
25. Kuo, R. J., Lin, S. Y., & Shih, C. W. (2007). Mining Association Rules through Integration of Clustering Analysis and Ant Colony System for Health Insurance Database in Taiwan. *Expert System with Application*, 33, 794-808.
26. Ravisankar, P., V. Ravi, G. Raghava Rao & I. Bose. (2011). Detection of Financial Statement Fraud and Feature Selection Using Data Mining Techniques. *Decision Support Systems*, 50, 491-500.
27. Songini, M. L. (2004). Fraud Sniffers. *Computer World*, 38.
28. Spathis, C. (2002). Detecting False Financial Statements Using Published Data: Some Evidence from Greece. *Managerial Auditing Journal* 17(4): 179-191.
29. Spathis, C., E. Kirkos & Y. Manolopoulos (2006). Data Mining Techniques for the Detection of Fraudulent Financial Statements. *Expert Systems with Applications* 32 (4): 995-1003.
30. Summers, S. L. & J. T. Sweeney. (1998). Fraudulently Misstated Financial Statements and insider Trading: An Empirical Analysis. *Accounting Review* 73 (1): 131- 146.
31. Wu, R., a, C.S. Ou, H. Lin, She-I. Chang & David C. Yen. (2012). Using Data Mining Technique to Enhance Tax Evasion Detection Performance. *Expert Systems with Applications* 39 8769-8777.