

کاربرد تکنیک‌های داده کاوی جهت نمونه‌گیری اظهارنامه‌های مالیاتی

حمید مستوفی

کارشناس ارشد حسابداری و کارشناس ارشد مالیاتی و مدرس دانشگاه پیام نور
همدان، hamidm_2004@yahoo.com

چکیده

سازمان امور مالیاتی جهت جلوگیری از فرار مالیاتی و اطمینان از تمکین مؤدیان مالیاتی، نیاز به حسابرسی تمام یا اکثر مؤدیان مالیاتی را دارد. اما این کار هم برای سازمان امور مالیاتی و هم برای مؤدیان مالیاتی زمانبر و پر هزینه است. حسابرسی هم برای مالیات دهندگان و هم برای اقتصاد فرایندی ناخوشایند است. بنابراین سازمان امور مالیاتی باید منابع محدود را به نحو معقول و منطقی تخصیص دهد تا به حداکثر تمکین مؤدیان مالیاتی با کمترین هزینه و خطا دست پیدا کنند. این مقاله به بررسی استفاده از الگوریتم‌های داده کاوی به‌عنوان گزینه‌ای با صرفه اقتصادی جهت نمونه‌گیری کارا و اثر بخش اظهارنامه‌های مالیاتی جهت حسابرسی مالیاتی می‌پردازد.

طبقه‌بندی JEL: H2

کلیدواژه‌ها: حسابرسی مالیاتی، تمکین مالیاتی، تکنیک‌های داده کاوی، استراتژی نمونه‌گیری حسابرسی

۱- مقدمه

در هر کشوری، مالیات به‌عنوان یکی از ابزارهای اصلی دولت در اقتصاد، نقش بسیار مهمی را ایفا می‌کند به گونه‌ای که امروزه در کشورهای صنعتی، بهبود و توسعه کارایی و اثربخشی اظهار مالیات مودیان و مطالبه آن توسط کارشناسان امور مالیاتی ضمن رعایت اصل عدالت مالیاتی، از اهمیت زیادی برخوردار گردیده است (پژویان، ۱۳۸۰) متناسب با توجه روزافزون به مقوله مالیات در سطح بین‌المللی، در کشور ایران نیز اقداماتی در این زمینه صورت پذیرفته است. تدوین طرح جامع مالیاتی و تاکید بر اجرایی کردن آن در سطح کشور، دلیل محکمی بر این ادعا است. پروژه انتخاب برای حسابرسی مبتنی بر ریسک به‌عنوان یکی از پروژه‌های اولویت دار سازمان امور مالیاتی کشور در طرح جامع مالیاتی می‌باشد شناسایی اظهارنامه‌های مالیاتی پر ریسک و قرار دادن آنها در نمونه انتخابی حسابرس مالیاتی، به منظور رسیدگی در هر نظام مالیاتی از اهمیت زیادی برخوردار است. در خصوص اهمیت انتخاب برای حسابرسی مبتنی بر ریسک لازم به ذکر است که در روال جاری، تمام اظهارنامه‌های مالیاتی رسیدگی می‌شوند و تنها در طرح خود اظهاری موضوع ماده 58 ق.م.م.د درصدی از اظهارنامه‌ها، به‌طور تصادفی انتخاب و مورد رسیدگی قرار می‌گیرند. این انتخاب تصادفی باعث می‌شود تا در برخی موارد، اظهارنامه‌های پر ریسک که از نظر میزان مالیات با اهمیت تلقی می‌شوند، در نمونه انتخابی کارشناسان امور مالیاتی به منظور رسیدگی به صحت و سقم مالیات ابرازی قرار نگیرند. نتایج این تحقیق می‌تواند در عملیاتی شدن پروژه مذکور مورد استفاده قرار گیرد.

۱-۱- حسابرسی مالیاتی

حسابرسی مالیاتی عبارتست از بررسی دقیق فعالیت‌های مؤدیان مالیاتی جهت اطمینان از اینکه آیا مودی مالیاتی بدهی مالیاتی خود را به نحو صحیحی ابراز نموده است یا خیر. حسابرسی مالیاتی نقشی اساسی در سازماندهی مالیاتی و دستیابی به درآمدهای مالیاتی مورد نظر دارد و موجب اطمینان از سلامت مالی کشور می‌شود و به‌عنوان یک عامل بازدارنده است که باعث افزایش صداقت مالیات دهندگان می‌شود اما به خاطر ماهیت مداخله‌جویانه، حسابرسی هم برای مالیات دهندگان و هم برای اقتصاد فرایندی ناخوشایند است. انجام حسابرسی هم برای مؤدیان و هم برای سازمان امور مالیاتی هزینه بر است بنابراین ادارات امور مالیاتی باید از حداقل منابع به نحو خردمندانه‌ای استفاده کنند تا به حداکثر تمکین مؤدیان دست پیدا کنند. بنابراین

استراتژی حسابرسی مناسب یک عامل کلیدی در کشف تقلب‌های صورت گرفته است. (شائو چانگ، ژائو، ۲۰۰۲)

نمونه‌گیری اظهارنامه‌های مالیاتی برای رسیدگی همانند گشتن دنبال سوزن در انبار کاه است. هر سال تعداد زیادی از مؤدیان مالیاتی درآمد خود را به درستی ابراز نمی‌کنند لذا سازمان امور مالیاتی ناچار است تا اقداماتی را جهت شناسایی و ملزم کردن آنها به تمکین انجام دهد، بدون اینکه افزایشی در هزینه تمکین ایجاد شود. اداره امور مالیاتی با حجم وسیعی از اطلاعات مؤدیان روبروست. غیرممکن است که با نگاهی ساده به اظهارنامه و پرونده مؤدیان مالیاتی به صحیح بودن و قصد و نیت آنان پی برد. با توجه به این واقعیت بهترین گزینه با صرفه، استفاده از الگوریتم‌های داده کاوی است. (جان ودیک، ۱۹۸۳)

برای سازمان امور مالیاتی ممکن نیست که یک روش نمونه‌گیری را از کشورهای دیگر اقتباس نموده و استفاده نماید زیرا در کشور هایی که از تکنیک انتخاب برای حسابرسی مبتنی بر ریسک در گزینش اظهارنامه‌ها به منظور حسابرسی مالیاتی استفاده می‌کنند، روش انتخاب شاخص‌ها و معیارها و نیز چگونگی طبقه‌بندی اظهارنامه‌ها بر مبنای ریسک، کاملاً محرمانه است. زیرا با افشاء نحوه انتخاب برای حسابرسی مبتنی بر ریسک، برخی اشخاص فرصت طلب (حقیقی و حقوقی) به منظور پایین آوردن ریسک انتخاب، اقدام به دستکاری اظهارنامه‌های مالیاتی خود می‌نمایند. از این رو، انتخاب برای حسابرسی مبتنی بر ریسک و نحوه عملکرد آن در هر کشوری کاملاً سری بوده و اقتباس چنین سیستمی از دیگر کشورها شدنی نیست. تکنیک‌های داده کاوی به صورت بسیار گسترده‌ای برای یادگیری الگوی داده‌های گذشته مورد استفاده قرار می‌گیرند و از طریق این الگوی یادگیری شده می‌توان رفتار داده‌های حال حاضر را پیش بینی نمود. در این شرایط، داده کاوی ابزاری سودمند است که در اختیار تمام شرکت‌ها می‌باشد. داده کاوی و کشف دانش، ابزارهایی هوشمند هستند که به جمع آوری و پردازش داده‌ها و بهره‌مندی از آنها، منجر می‌شوند. در واقع، داده کاوی شامل استفاده از ابزارهای تحلیلی پیچیده به منظور کشف الگوهای معتبر و ناشناخته و وابستگی های موجود در مجموعه داده‌های حجیم می‌باشد. نتیجه این که، فناوری داده کاوی چیزی بیش از جمع‌آوری و مدیریت داده بوده و شامل تجزیه و تحلیل و پیش‌بینی نیز می‌باشد. (روزاریو ماناسان، ۲۰۰۳)

وجود نارسایی در قوانین موجود، کمبود ممیزان و مامورین مالیاتی مجرب، نزولی بودن نرخ‌های مالیاتی در عمل، برخوردهای ناعادلانه مالیاتی، غیرعادلانه بودن نرخ‌های مالیاتی، کندی سیستم مالیاتی، تبنانی مودی و ممیز، حساب‌سازی‌های مالیاتی و فقدان یک مرکز قوی اطلاعات مالیاتی از عوامل بروز فساد مالی و اقتصادی در حوزه مالیاتی می‌باشد. این موضوعات باعث می‌شود که هر ساله درصدی از تولید ناخالص داخلی کشور خارج از شمول مالیات قرار گرفته و موجب عدم تحقق اهداف پیش‌بینی شده در مورد درآمدهای مالیاتی و در نتیجه عدم تحقق عدالت مالیاتی شود. همان‌طور که ذکر شد، عمده عوامل بروز فساد مالی و اقتصادی در حوزه مالیاتی عمدتاً مربوط به دو حوزه عالی (سیاست‌گذاری و قانون‌گذاری) و عملیاتی (رسیدگی و ممیزی) می‌باشند. (هو دانیل، لاو، ۱۹۹۹)

داده کاوی می‌تواند عوامل مؤثر بر گریز مالیاتی را شناسایی و مدل‌هایی را جهت کشف میزان احتمالی گریز مالیاتی مودیان ارائه دهد. بنابراین، دانش ایجاد شده از فرآیند داده کاوی راهکارهایی را به سیاست‌گذاران حوزه مالیاتی در خصوص قانون‌گذاری و چارچوبی را به ممیزان مالیاتی جهت رسیدگی کارا و اثربخش ارائه می‌دهد. در حوزه عملیاتی (رسیدگی و ممیزی)، داده کاوی پا را فراتر گذاشته و می‌تواند مبنایی جهت پیاده‌سازی سیستم حسابرسی مبتنی بر ریسک در سازمان امور مالیاتی فراهم آورد. بر این اساس ممیزان مالیاتی می‌توانند قبل از نمونه‌گیری و انجام رسیدگی، کلیه مودیان مالیاتی را در بازه‌ای از ریسک پایین تا بالا (به لحاظ احتمال گریز مالیاتی) دسته‌بندی و برنامه‌های رسیدگی خود را بر اساس این دسته‌بندی تدوین و اجرا نمایند. به این ترتیب ممیزان مالیاتی می‌توانند در خصوص این سوال که کدامیک از مودیان نیاز به بررسی بیش‌تری دارند، تصمیمات مناسبی را اتخاذ نمایند. این رویه از یک طرف موجب افزایش کارایی و اثربخشی رسیدگی‌ها شده و از طرف دیگر موجب کاهش هزینه‌های رسیدگی و افزایش درآمدهای مالیاتی سازمان امور مالیاتی می‌شود.

۲- مرور تحقیقات پیشین

الم بلاک می و مک کی ۲۰۰۴ روش نمونه‌گیری حسابرسی مالیات بر فروش در آمریکا و تأثیر آن بر تمکین مالیاتی را بررسی نمودند. نتایج نشان داد که ممیزان مالیاتی اظهارنامه‌ها را بر اساس روش سیستماتیک جهت حسابرسی انتخاب می‌کنند و

شرکت‌هایی که بیش‌ترین انحراف را دارند و در آخرین لحظه اظهارنامه خود را می‌دهند و آدرس درستی ندارند نرخ تمکین پایین‌تری دارند.

مورای ۱۹۹۵ به بررسی استراتژی نمونه‌گیری برای حسابرسی مالیات بر فروش و دلایل پائین‌تر از واقع ابراز نمودن فروش توسط مؤدیان پرداخت. نتایج نشان داد که حساب مالیات بر فروش به صورت تصادفی برای حسابرسی انتخاب نمی‌شود. هم‌چنین نتایج نشان داد که کم‌تر ابراز نمودن فروش با رفتارهای ناظر بر شرکت همبستگی دارد.

کومار ۱۹۹۹ یک سیستم نمونه‌گیری بر مبنای معیار را در هند ارائه نمود و نسبت‌های مختلفی را در ترازنامه مورد آزمون قرار داد مثل رشد نسبت گردش وجودی، نسبت جاری، نسبت دارایی به گردش موجودی و ...

فوا، الاهاکون و لی ۲۰۰۵ در مقاله خود روش‌ها و تکنیک‌های کشف تقلب منتشر شده در تحقیقات علمی و صنعتی انجام شده در ۱۰ سال گذشته را گروه بندی، مقایسه و تلخیص نمودند.

۳- اهداف

هدف این مقاله ارائه مدلی برای نمونه‌گیری اظهارنامه‌های مالیاتی جهت حسابرسی مالیاتی توسط سازمان امور مالیاتی کشور است.

۴- روش کار و توصیف داده‌ها

مؤدیان مالیاتی به دو گروه بر اساس اطلاعات پیشین تقسیم شدند. گروه اول مودیانی بودند که متوجه شدیم درآمد خود را کم‌تر از واقع ابراز نموده بودند یا فرار مالیاتی داشتند. ۴۰۲ مودی انتخاب شد.

نمونه دوم را می‌توان مؤدیان خوب نامید و با اطمینان زیاد می‌توان گفت که مالیات خود را به درستی ابراز کرده‌اند. برای نمونه دوم که مالیات زیادی را پرداخت نموده‌اند فرض می‌شود که مالیات خود را به درستی ابراز کرده‌اند. با این حال باید احتیاط لازم را به‌کاربرد تا نمونه به‌خوبی بیانگر جامعه مورد نظر باشد به همین خاطر کالاهای عمده و ماهیت تجارت (خرده فروش / عمده فرش / صادرکننده) مؤدیان شناسایی شد. سپس برای هریک از مؤدیان گروه اول (فراریان) سه نفر از عمده‌ترین پرداخت کنندگان مالیات در هر گروه کالا و ماهیت تجاری شناسایی شدند. لیست آزمایشی مؤدیان خوب تهیه شد. فرایند ذکر شده حدود ۱۲۰۰ مودی را تعیین می‌نماید.

بعد از دسته‌بندی مؤدیان به دو گروه، روش مناسب داده کاوی انتخاب شد و پارامترهای مهم در فرار مالیاتی و کم‌تر از واقع گزارش کردن درآمدها مشخص شد. به منظور شناسایی و ارائه مدل مناسب از سه روش داده کاوی استفاده نمودیم: درخت طبقه‌بندی، رگرسیون لجستیک و تابع تشخیص با توجه به تحقیقات انجام شده پیشین، توان ممیزان مالیاتی و اینکه آیا داده‌ها را می‌توان از اظهارنامه و پایگاه داده استخراج کرد پارامترهای زیر برای داده کاوی انتخاب نمودیم.

الف. مشخصات مودی

- الف ۱. ثبت نام کننده جدید (بلی/خیر)
- الف ۲. ضریب مالیاتی بالا (بلی/خیر)
- الف ۳. فعالیت‌های تجاری دیگری در همان آدرس (بلی/خیر)
- الف ۴. مؤسسه بازرگانی دیگری با همان شماره تلفن (بلی/خیر)

ب. تمکین از طریق اظهارنامه

- ب ۱. هرگونه جای خالی در اظهارنامه (بلی/خیر)
- ب ۲. تأخیر در پر کردن اظهار نامه (تعداد روزها)
- ب ۳. تعداد اظهارنامه‌هایی که خالی است یا مخدوش شده

ج. نسبت‌ها و مقادیر اظهارنامه

- ج ۱. نسبت مالیات بر گردش موجودی
- ج ۲. درصد سود ناخالص
- ج ۳. فروش معاف بر گردش موجودی
- ج ۴. نسبت موجودی بر گردش موجودی
- ج ۵. نسبت خرید به فروش
- ج ۶. اعتبار مالیاتی منتقل شده بیش‌تر از ۱۰۰۰۰۰۰۰ ریال است (بلی/خیر)
- ج ۷. تعدیلات مالیاتی خروجی یا ورودی بیش‌تر از ۱۰۰۰۰۰۰۰ ریال است (بلی/خیر)

ح. انحراف (مغایرت) در اظهارنامه‌های مالیاتی ارائه شده در دوره‌های

مختلف مالیاتی

- ح ۱. رشد گردش موجودی در مقایسه با پارسال
- ح ۲. رشد مالیات در مقایسه با پارسال
- ح ۳. انحراف در گردش موجودی در طول دوره‌های مختلف مالیاتی

- ح ۴. تغییر/انحراف در فروش کل (ملی/داخل استانی/صادراتی)
 ح ۵. تغییر/انحراف در خرید کل (ملی/داخل استانی/صادراتی)
 ح ۶. تغییر/انحراف در تولیدات برای فروش
 ح ۷. تغییر/انحراف در تولیدات برای خرید
 ج. مقایسه فروشندگان یک صنعت/صنف خاص با توجه به پارامترهای زیر:
 ج ۱. مالیات به گردش موجودی
 ج ۲. درصد سود ناویژه
 ج ۳. فروش مشمول مالیات به گردش موجودی
 ج ۴. موجودی کالا به گردش موجودی
 ج ۵. افزایش گردش موجودی
 ج ۶. افزایش مالیات

متغیر هدف این است که آیا فروشندگان مالیاتشان را کمتر از واقع گزارش کرده‌اند یا خیر. دو گزینه برای تعریف متغیر هدف وجود دارد. الف) مقدار مالیات کمتر از واقع گزارش شده که کشف شده است ب) یا این حقیقت که کمتر از واقع گزارش کردن مالیات کشف شده است (بلی/خیر). استراتژی حسابرسی مالیاتی افزایش درآمدهای مالیاتی نیست بلکه افزایش نرخ کشف حسابرسی است و موجب ترس از حسابرسی مؤدیان می‌گردد. متغیر هدف را برای مؤدیان شناسایی شده‌ای که مالیات خود را کمتر ابراز نموده‌اند "۱" قرار می‌دهیم و برای مؤدیان که بزرگ‌ترین پرداخت کنندگان مالیات در گروه خود هستند و هیچ موردی در مورد آنها کشف نشده است "۰" قرار می‌دهیم.

۵- تجزیه تحلیل و نتایج

۵-۱- روش ارزیابی

هیچ استراتژی حسابرسی نمی‌تواند کل فرار مالیاتی را شناسایی نماید. سازمان امور مالیاتی تمایل دارد که منابع خود را افزایش دهد که این امر موجب افزایش ناراضی (نرخ شناسایی) می‌گردد اما ترس از حسابرسی به صورت اتوماتیک سبب افزایش درآمدهای مالیاتی گردد. بنابراین هدف یک استراتژی حسابرسی باید شناسایی مؤدیان باشد که درآمد و مالیات خود را کمتر از واقع ابراز نموده‌اند. با استفاده از این روش، ترس از حسابرسی به‌طور اتوماتیک باعث افزایش درستکاری مؤدیان و در نتیجه افزایش درآمدهای مالیاتی می‌گردد.

بر اساس این واقعیت که مؤدیان مالیاتی فرار می‌کنند یا خیر و مقایسه آن با پیش بینی انجام شده، چهار حالت ممکن که به‌عنوان ماتریس پچیدگی شناخته می‌شود وجود خواهد داشت.

جدول ۱- ماتریس پچیدگی

۱ فرار مالیاتی	۰ تمکین مالیاتی	پیش بینی شده ←
		واقعی ↓
مثبت کاذب	منفی واقعی	۰ تمکین مالیاتی
مثبت واقعی	منفی کاذب	۱ فرار مالیاتی

هدف مطلوب حسابرسی در فرمول‌های زیر نشان داده شده است. مثبت کاذب زمانی رخ می‌دهد که مدل، مودی مالیاتی را که تمکین کرده به‌عنوان فرار کننده از مالیات شناسایی کند و منفی کاذب نیز هنگامی رخ می‌دهد که مدل، مودی مالیاتی را که فرار مالیاتی دارد را به‌عنوان تمکین کننده شناسایی کند. درصد فرارهای مالیاتی که توسط مدلبه‌طور صحیح شناسایی گردیده است، کارایی پیش بینی (PE) نام دارد

کارایی پیش بینی (PE) = مثبت واقعی / (مثبت واقعی + منفی کاذب)

کاهش در تعداد مواردی که باید آزمون شوند در مقایسه با روش سنتی که تمام موارد باید آزمون گردند کاهش در تلاش آزمون (EF) نام دارد

$$EF = \frac{1 - (\text{منفی واقعی} + \text{مثبت واقعی})}{\text{کل موارد}}$$

درصد مواردی که احتمالاً فرار مالیاتی شناسایی می‌گردد، اگر موارد پیش بینی

شده حسابرسی گردند، نرخ شناسایی (SR) نام دارد

$$SR = \frac{\text{مثبت واقعی}}{\text{مثبت واقعی} + \text{منفی واقعی}}$$

1- Prediction efficiency

2- Reduction in examination effort

3- Strike Rate

هنگامی که هدف اصلی استفاده کارا از منابع باشد، مهم‌ترین پارامتر برای ارزیابی مدل، نرخ شناسایی است. البته کارایی پیش بینی را نمی‌توان نادیده گرفت زیرا اگر کارایی پیش بینی مدل پایین باشد، مدل غیرقابل استفاده می‌گردد.

۵-۲- ساختار مدل

مدل‌های مختلف داده کاوی مورد استفاده قرار گرفته که در این قسمت به شرح آنها می‌پردازیم:

مدل ۱: ابتدا از مدل درخت طبقه‌بندی برای دسته‌بندی نمونه‌ها به دو مجموعه آموزشی و مجموعه آزمایشی استفاده شد. در ابتدا مدل با استفاده از مجموعه آموزشی، تعیین و سپس با استفاده از مجموعه آزمایشی دقت آن امتحان می‌شود. داده‌ها در مجموعه‌های دوتایی قرار داده شدند و از داده‌های پردازش نشده استفاده شد. عملکرد مدل‌ها در جدول زیر نشان داده شده است.

جدول ۲- ارزیابی پارامترها در مدل ۱ تا ۶

پارامتر	کارایی پیش بینی	کاهش در تلاش آزمون	نرخ شناسایی
داده‌های آموزش داده شده	٪۳۷	٪۸۹	٪۹۳
داده‌های مورد آزمون	٪۳۱	٪۸۹	٪۷۵
مدل ۲			
داده‌های آموزش داده شده	٪۴۷	٪۸۸	٪۸۴
داده‌های مورد آزمون	٪۴۲	٪۸۵	٪۶۹
مدل ۳			
نتایج مشابه مدل ۲			
مدل ۴			
داده‌های کامل	٪۳۲	٪۹۰	٪۷۹
مدل ۵			
داده‌های اصلی	٪۵۹	٪۷۲	٪۵۱
داده‌های ارزیابی شده	٪۵۶	٪۷۱	٪۴۸
مدل ۶			
انتخاب فقط ۲	٪۳۰	٪۸۲	٪۴۰
انتخاب ۱ و ۲	٪۶۸	٪۶۳	٪۴۵

مدل ۱ دارای قدرت پیش بینی ضعیفی است و تنها می‌تواند ۳۱٪ داده‌ها را به صورت درست پیش بینی کند.

مدل ۲: برای افزایش قدرت پیش بینی، یک مدل درخت طبقه‌بندی دیگری طراحی شد. این بار داده‌های پردازش شده مورد استفاده قرار گرفت. متغیرهای ورودی به ۴ گروه با عمق برابر تقسیم شدند. نتایج در جدول ۲ نشان داده شده است.

کارایی پیش بینی در این مدل به ۴۱-۴۲٪ ارتقا یافته است. هم‌چنین نرخ شناسایی ۷۰٪ برای یک استراتژی حسابرسی اثربخش مناسب به نظر می‌رسد.

مدل ۳: درخت طبقه‌بندی دیگری برای تجدید طبقه‌بندی داده‌ها طراحی گردید. "۰" برای تمکین مالیاتی و "۱" برای عدم تمکین مالیاتی. داده‌ها دوباره طبقه‌بندی گردید. اما این مدل هم نتیجه‌ای بهتر از مدل ۲ نداشت لذا کنار گذاشته شد.

مدل ۴ و ۵: با توجه به اینکه بهبودی در قدرت پیش بینی مدل حاصل نشد و توجه به این امر که بسیاری از داده‌های ورودی به صورت نسبت هستند، لذا از روش رگرسیون لجستیک و تابع تشخیص استفاده شد که عملکرد آنها در جدول دو نمایش داده شده است.

مدل رگرسیون لجستیک عملکرد بهتری از مدل ۲ نداشته و کارایی پیش بینی و نرخ شناسایی هر دو مدل پائین است. مدل تابع تشخیص کارایی پیش بینی را تا ۵۷٪ بهبود می‌بخشد هر چند این امر با کاهش نرخ شناسایی به حدود ۵۰٪ همراه است. مدل مناسب است اما نرخ شناسایی آن پائین است.

مدل ۶: مدل تابع تشخیص باعث بهبود کارایی پیش‌بینی شد اما این امر با افزایش تلاش‌های آزمون میسر گردیده است. مدل مورد استفاده بعدی مدل تابع تشخیص است اما این بار متغیر هدف در سه سطح همانند مدل ۳ کد گذاری شده است. نتایج در جدول ۲ نشان داده شده است. این مدل تابع تشخیص باعث افزایش کارایی پیش بینی به ۶۸٪ می‌شود هر چند باعث کاهش نرخ شناسایی و افزایش تلاش‌های آزمون می‌گردد. مدل بسیار مناسب است اما نرخ شناسایی را کاهش می‌دهد.

مدل ۷: مدل تابع تشخیص در کارایی پیش بینی و مدل درخت تصمیم‌گیری در نرخ شناسایی بهتر هستند به منظور بهره مندی از این دو مزیت، یک مدل هیبریدی طراحی گردید. داده‌ها ابتدا از طریق مدل تابع تشخیص (مدل ۶) مورد پردازش قرار گرفته و سپس خروجی این مدل در مدل درخت طبقه‌بندی قرار داده می‌شود. این کار

به دو روش انجام می‌گیرد. روش اول این است که درخت طبقه‌بندی مستقلی را برای آزمون و آموزش داده‌ها استفاده نمود و روش دوم اینکه درخت طبقه‌بندی مهیا شده مدل ۲ برای خروجی تابع تشخیص مورد استفاده قرار گیرد. ابتدا یک درخت طبقه‌بندی مستقلی انتخاب شد (مدل ۶)

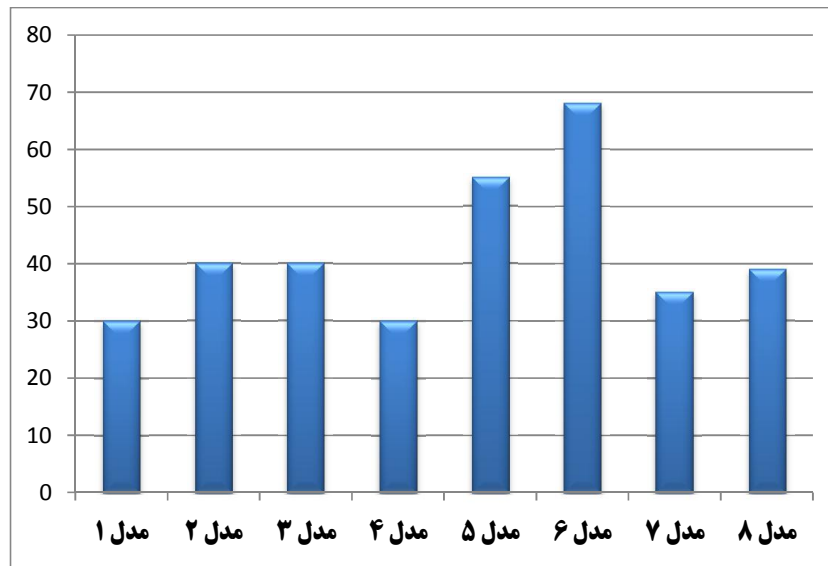
جدول ۳- پارامترهای ارزیابی مدل ۷ و ۸

پارامتر	کارایی پیش‌بینی	کاهش در تلاش آزمون	نرخ شناسایی
مدل ۷			
خارج از داده‌های انتخابی	۸۲	۵۵	۵۴
مقایسه با کل داده‌ها	۳۶	۸۳	۵۴
مدل ۸			
خارج از داده‌های انتخابی	۵۸	۷۰	۸۶
مقایسه با کل داده‌ها	۳۹	۸۹	۸۶

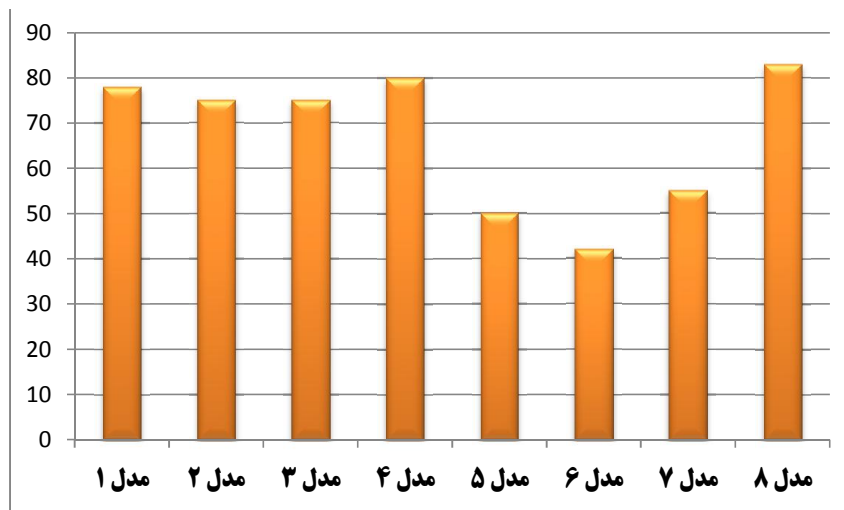
همان‌طور که مشخص است این مدل حدود ۸۲٪ موارد را به درستی پیش‌بینی می‌کند اما نرخ شناسایی پایین است به خاطر تعداد زیادی مثبت کاذب که مدل آنها را شناسایی نکرده است.

مدل ۸: درخت طبقه‌بندی در مدل هیبریدی به خاطر اینکه مثبت کاذب را شناسایی نمی‌کند، مناسب نیست. هم‌چنین درخت طبقه‌بندی مدل ۲ دارای بالاترین نرخ شناسایی می‌باشد. لذا درخت طبقه‌بندی مدل ۲ را برای طبقه‌بندی داده‌های خروجی از تابع تشخیص مورد استفاده قرار می‌دهیم. این مدل هیبریدی دارای نرخ شناسایی بالا و میزان تلاش آزمون پائین است.

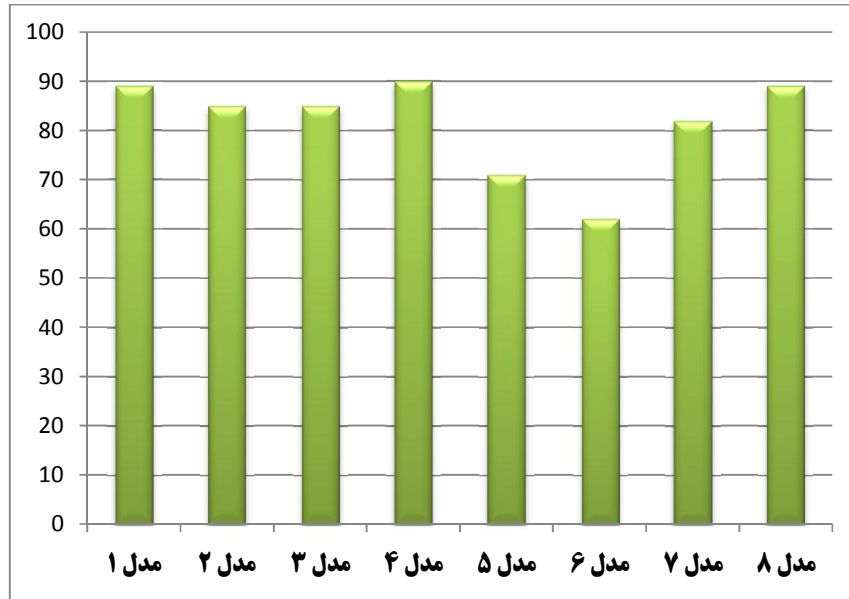
همان‌طور که در ادامه خواهیم گفت برای هدف استراتژی حسابرسی مالیاتی، باید مدل دارای یک نرخ کارایی پیش‌بینی منطقی و یک نرخ شناسایی خوب باشد. شکل ۱ زیر به صورت مقایسه‌ای پارامترهای هر مدل را نشان می‌دهد. مقایسه کارایی پیش‌بینی مدل با روش انتخاب تصادفی اظهارنامه‌ها جالب است. همه مدل‌ها نرخ شناسایی بهتری از انتخاب تصادفی دارند.



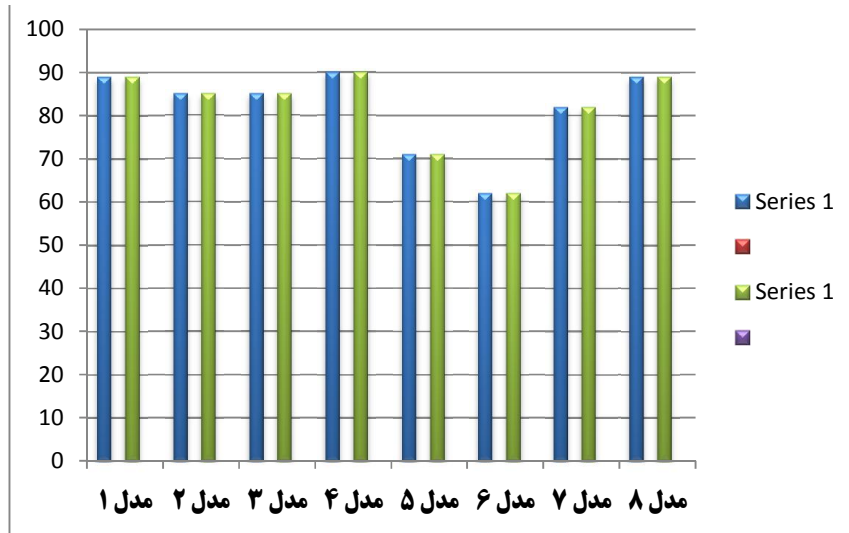
شکل ۱- کارایی عملکرد مدل‌های مختلف



شکل ۲- نرخ شناسایی مدل‌هایی مختلف



شکل ۳- کاهش تلاش آزمون در هر مدل



شکل ۴- مقایسه کارایی مدل‌ها با کارایی روش انتخاب تصادفی

۶- نتیجه‌گیری

نوعی موازنه یا مصالحه بین نرخ شناسایی و کارایی پیش بینی وجود دارد که مدل‌ها به صورت همزمان نمی‌توانند به بهبود هر دو مورد کمک کنند. همان‌طور که قبلاً ذکر شد هدف حسابرسی کشف همه فراریان مالیاتی نیست بلکه ترس از حسابرسی شدن خود باعث افزایش درآمدهای مالیاتی می‌گردد. بنابراین در مدل انتخاب مودیان، ممکن است بین نرخ پیش بینی و نرخ شناسایی مصالحه انجام شود. مدل ۲ (درخت طبقه‌بندی) و مدل ۸ (هیبرید) هر دو دارای نرخ شناسایی خوب (۷۰٪ و ۸۶٪) و نرخ کارایی پیش بینی معقول (۴۱٪ و ۳۹٪) می‌باشند. تمام مدل‌های مبتنی بر داده کاوی از روش انتخاب تصادفی بهتر بودند. نرخ شناسایی در مقایسه با مدل تصادفی به میزان دو برابر در مدل تابع تشخیص، ۲.۵ برابر در مدل درخت طبقه‌بندی، سه برابر در مدل رگرسیون لجستیک و ۳.۵ برابر در مدل هیبریدی بهتر می‌باشد.

فهرست منابع

- پژویان، جمشید (۱۳۸۰)، اقتصاد بخش عمومی مالیات‌ها، تهران: دانشگاه تربیت مدرس.
- Alm, James, Blackwell C. & McKee Michael. (2004) Audit Selection and Firm Compliance with a Broad-based Sales Tax. *National Tax Journal*, 57, (2), 1, 209-27.
- Ho, Daniel, Peter Lau. (1999) Tax Audits in Hong Kong. *The International Tax Journal New York*, 25, (3), 61-71.
- Kumar, Anuj and V. Nagadevara. (2006) Development of Hybrid Classification Methodology for Mining Skewed Data Sets – A Case Study of Indian Customs Data, *Proceedings of the 4th ACS/IEEE International Conference on Computer Systems and Applications*, Mar 8-10, 2006, Sharjah, UAE.
- Manasan Rosario G. (2003) Estimating Industry Benchmarks for the Value Added Tax. *Philippine Journal of Development*, 55, (30), 1, 71-90.
- Murray, Mathew N. (1995) Sales Tax Compliance and Audit Selection. *National Tax Journal*. 48, (4), 515-30.
- Wedick, John L. (1983) Looking for a Needle in a Haystack – How the IRS Selects returns for Audit. *The Tax Advisor*, New York, 14, (11), 675-675.
- Phua C, Alahakoon D, Lee V. Minority Report in Fraud Detection: Classification of Skewed Data, *SIGKDD Explorations*, Vol 6, (1), 50-59.

Shao H, Zhao H, Chang G. (2002) Applying Data Mining to Detect Fraud Behavior in Customs Declaration, *Proceedings of the International Conference on Machine Learning and Cybernetics*, Vol 3, IEEE 2002, 1241-1244.