

مقایسه درخت تصمیم (الگوریتم C5.0 و جنگل تصادفی) و ماشین بردار پشتیبان در اعتبارسنجی مؤدیان مالیاتی

هوشنگ خرم‌نیا^۱

میرفیض فلاح شمس^۲

غلامرضا زمردیان^۳

علی اصغر انواری رستمی^۴

چکیده

امروزه روش حسابرسی مبتنی بر ریسک در نظام‌های نوین مالیاتی مورد تأکید است، لذا تبیین مدلی جامع جهت رتبه‌بندی ریسک مؤدیان یکی از گام‌های اساسی اجرای طرح جامع مالیاتی است. بنابراین در این مقاله برآنیم تا عملکرد الگوریتم‌های درخت تصمیم و ماشین بردار پشتیبان را در اعتبارسنجی مؤدیان مالیاتی مقایسه کنیم. جامعه آماری این پژوهش شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران هستند که طی سالهای ۱۳۹۲ تا ۱۳۹۷ فعال بوده‌اند و برای انتخاب نمونه از روش غربالگری (حذفی) استفاده شده‌است. در این پژوهش ابتدا با استفاده از تکنیک دلفی و متاستنز ۱۶۴ مؤلفه مؤثر در اعتبارسنجی مؤدیان مالیاتی شناسایی گردید. سپس داده‌های مورد نیاز برای اندازه‌گیری متغیرهای پژوهش از سایت کدال و با بررسی پرونده‌های مالیاتی استخراج شد و در نهایت با استفاده از داده‌های جمع‌آوری شده به بررسی میزان دقت درخت تصمیم (الگوریتم C5.0 و جنگل تصادفی) و ماشین بردار پشتیبان در اعتبارسنجی مؤدیان مالیاتی پرداختیم. یافته‌ها نشان داد براساس نتایج حاصل از مقدار AUC^5 الگوریتم C5.0 و جنگل تصادفی برآزش بهتری دارند با این وجود فرضیه پژوهش مبنی بر امکان پیش‌بینی ریسک مؤدیان مالیاتی با استفاده از الگوریتم SVM رد نمی‌شود.

واژه‌های کلیدی: اعتبارسنجی مؤدیان، محافظه‌کاری، جریان وجه نقد عملیاتی، تمکین مالیاتی

تاریخ دریافت: ۱۴۰۲/۰۶/۰۴، تاریخ پذیرش: ۱۴۰۲/۰۹/۰۸

۱. دانشجوی دکتری، گروه مدیریت مالی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران، khoramniah@gmail.com

۲. دانشیار گروه مدیریت مالی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران، (نویسنده مسئول)، fallahshams@gmail.com

۳. دانشیار گروه مدیریت مالی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران، gh.zomorodian@gmail.com

۴. استاد گروه حسابداری، دانشگاه تربیت مدرس، تهران، ایران، anvary@modares.ac.ir

مقدمه

سازمان امور مالیاتی برای اصلاح نظام مالیاتی با رویکرد تحقق عدالت مالیاتی، تسریع و تسهیل در انجام امور مالیاتی و ارتقاء سطح رضایتمندی مؤدیان مالیاتی و شفاف‌سازی فعالیت‌های اقتصادی، برنامه‌های متعددی را در دستور کار قرار داده است. یکی از این آنها بهره‌گیری از حسابرسی‌های مدرن با رویکرد مدیریت ریسک بوده و این رویکرد، عنصری اساسی جهت مدیریت کارآ و اثربخش رعایت تمکین مالیاتی بحساب می‌آید. این دیدگاه می‌کوشد تا با اولویت‌بندی عملیات در سازمان مالیاتی، مسیر تعامل بین مؤدیان و سازمان را به سمت تبعیت مالیاتی بهتر از سوی مؤدیان و کاهش تنش‌های مالیاتی سوق دهد. با توجه به مشکلاتی از قبیل حجم انبوه و رو به افزایش اظهارنامه‌های مالیاتی و هزینه‌های بالای رسیدگی به تک تک آنها، محدودیت زمانی جهت رسیدگی به اظهارنامه‌ها، نبود معیارهای مدون جهت رسیدگی، محدودیت نیروی انسانی، وجود قضاوت سلیقه‌ای گسترده در تشخیص مالیاتی و عدم ارائه اظهارنامه توسط برخی از مؤدیان در سیستم مالیاتی کشور نیاز است تا رویکرد جدیدی در راستای بر طرف ساختن مسائل فوق اتخاذ گردد. فعالیت‌های فرار مالیاتی، همواره از مسائل مهم و مسأله ساز در هر کشوری بوده است.

بسیاری از سازمان‌های مالیاتی استراتژیهای حسابرسی خود را بر پایه روش‌های نمره دهی بر مبنای ریسک توسعه داده‌اند. به دلیل محدود بودن منابع دولت و تعداد بسیار زیاد مؤدیان، سازمان مالیاتی برای شناسایی و انتخاب مؤدی جهت حسابرسی، نیازمند روشی سیستماتیک و بر مبنای ریسک می‌باشد (OECD, 2004). هدف اصلی این تحقیق در مرحله اول شناسایی عوامل مؤثر در اعتبارسنجی مؤدیان مالیاتی و در مرحله دوم مقایسه درخت تصمیم (الگوریتم C5.0 و جنگل تصادفی) و ماشین بردار پشتیبان در اعتبارسنجی مؤدیان مالیاتی است. در ادامه این مقاله ابتدا مبانی نظری و پیشینه پژوهش، سپس فرضیه‌ها، روش‌شناسی، متغیرها و آزمون فرضیه‌ها ارائه می‌شود و در انتهای مقاله نیز بحث و نتیجه‌گیری و پیشنهادها پژوهش ارائه خواهد شد.

مبانی نظری و پیشینه پژوهش

امروزه درگفتمان اداری و حتی سیاسی سازمان‌های مالیاتی مدرن مؤدی‌مدار که بر پایه احترام به مؤدیان مالیاتی استوارند به جای "فرار مالیاتی" اغلب از "عدم تمکین" و یا "شکاف مالیاتی" که همزمان دو مشکل اساسی تمام نظام‌های مالیاتی یعنی فرار و اجتناب مالیاتی را پوشش می‌دهد، استفاده می‌شود. فرار و اجتناب مالیاتی به عنوان یک «پسماند» یا به تعبیر دقیق‌تر «نشتی»^۱ در سیستم‌های مالیاتی تمام کشورها وجود دارد و تفاوت بین این کشورها، تنها در اندازه فرار و اجتناب به عنوان درصدی از تولید ناخالص داخلی است. اگرچه، حذف کامل پدیده

1. Attrition

عدم تمکین در نظام مالیاتی، تقریباً غیر ممکن است، اما بسیاری از کشورهای مختلف جهان بویژه کشورهای توسعه یافته، با آگاهی کامل از این موضوع و به منظور کاهش اندازه شکاف یا عدم تمکین مالیاتی، به طور مستمر نسبت به طراحی و اجرای برنامه‌های مربوط با عناوین مختلف از جمله «برنامه کاهش شکاف مالیاتی» و یا «برنامه مدیریت ریسک تمکین» اقدام می‌کنند که به لحاظ روش‌شناسی و حتی جزئیات، کاملاً شبیه هم هستند. در حال حاضر، برنامه مدیریت ریسک تمکین، به عنوان یک رویکرد مستقل مدیریتی و یا به عنوان بخشی از برنامه راهبردی سازمان‌های مالیاتی کشورهای توسعه یافته، اجرا می‌شود و بهبود مستمر تمکین از طریق اجرای برنامه‌های مدیریت ریسک تمکین، به یکی از ارکان اصلی مدیریت مالیاتی در نظام‌های مالیاتی کشورهای توسعه یافته تبدیل شده است. شناسایی ریسک‌های مهم تمکین، ارزیابی و اولویت‌بندی این ریسک‌ها، تحلیل رفتار تمکین مؤدیان و طراحی و پیاده‌سازی استراتژی‌های برخورد با این ریسک‌ها از ارکان اصلی برنامه مدیریت ریسک تمکین می‌باشد، از این رو اولین قدم برای مبارزه با فرار و اجتناب اجرای برنامه مدیریت ریسک تمکین و ایجاد ساختار سازمانی مناسب برای آن می‌باشد (Iranian National Tax Administrations Research Database, 1401).

مدیریت ریسک‌های تمکین، یک فرآیند مستمر چند مرحله‌ای از بالا به پایین است و هدف از آن شناسایی و رفع ریسک‌های سطوح مختلف و همچنین شناسایی و برخورد با مؤدیان پرریسکی است که تحت هیچ شرایطی حاضر به تمکین نیستند. پیش‌نیازهای سیستم مدیریت ریسک تمکین مالیاتی، اصلاح ساختار سازمانی با محوریت مدیریت تمکین، گروه‌بندی مؤدیان براساس برخی ویژگی‌های مشترک، ایجاد موتور ریسک، تقویت پایگاه داده، استفاده از فنون تحلیلی بویژه علوم داده برای تحلیل رفتار مؤدیان، آموزش سطوح مختلف سازمان با مفاهیم و فنون مدیریت ریسک تمکین مالیاتی و سرانجام طراحی برنامه مدیریت تمکین می‌باشد. بعد از برطرف کردن پیش‌نیازها، اولین مرحله از فرآیند مدیریت ریسک تمکین مالیاتی، شناسایی ریسک‌های تمکین است. به طور کلی، ریسک‌های تمکین مالیاتی در چند سطح از بالا به پایین پراکنده شده‌اند و در مرحله اول باید به شناسایی و رفع ریسک‌های سطح بالا یا کلان (یا دائمی) اقدام شود. در این سطح، دوگروه عمده از ریسک‌های موجود در قوانین و مقررات مالیاتی و ریسک‌های موجود در فرآیندهای مالیاتی وجود دارد. بعد از شناسایی و رفع ریسک‌های کلان، سازمان باید به سراغ ریسک‌های موجود در سطوح پایین‌تر یعنی سطح صنعت / فعالیت و سطح مالیات‌ها برود. در این مرحله باید به طور مستمر نسبت به شناسایی و تحلیل ریسک‌های این سطوح اقدام شود. آخرین مرحله از فرآیند مدیریت تمکین، شناسایی ریسک‌های تمکین در سطح مؤدی یا سطح پرونده است. در این سطح، شاخص‌های رفتاری تک تک مؤدیان در موتور ریسک تحلیل، ارزیابی و در نهایت ریسک هر مؤدی محاسبه و بر اساس آن، نسبت به انتخاب پرونده‌ها برای حسابرسی مالیاتی، اقدام می‌شود. ریسک‌های سطح مؤدی را نباید تنها با اتکا به شاخص‌های عمدتاً مالی بررسی و محاسبه کرد. مدیریت تمکین، یک فرآیند چند مرحله‌ای مستمر

شامل شناسایی، تحلیل، ارزیابی و رفع تمامی ریسک‌های تمکین است و رعایت این توالی، لازمه اجرای موفق این برنامه و ارتقای تمکین در سازمان است و شناسایی و انتخاب مؤدیان پریسک برای انجام حسابرسی، آخرین مرحله از فرآیند مدیریت ریسک تمکین است. فراموش نکنیم که سازمان با انسان‌ها و رفتارهای انسانی سروکار دارد و «تمکین» یک مقوله رفتاری انسان محور است و لذا نگاه مکانیکی به آن نمی‌تواند به بهبود شاخص تمکین کمک نماید و اصولاً با زدن چندکلیک، امکان مدیریت و کاهش ریسک تمکین مالیاتی وجود ندارد. این سیستم‌ها صرفاً پشتیبان تصمیم‌گیران هستند و نیروی انسانی توانمند و متخصص در زمینه علوم داده، تحلیل ریسک، اقتصاد رفتاری، آمار و حسابرسی، کامپیوتر و حتی روانشناسی مهمترین مؤلفه‌های یک سیستم کارآمد مدیریت ریسک تمکین می‌باشند (Iranian National Tax Administrations Research Database, 1401).

در یک سیستم مدرن جهت وصول مالیات غیرممکن است دولت تک تک مؤدیان را مورد ممیزی و رسیدگی قرار دهد و بررسی مداوم مؤدیان کم خطر و تمکین کننده باعث هدر رفتن منابع دولت می‌شود (Askari et al, 1394). در چنین سیستمی انتخاب مؤدی جهت حسابرسی باتوجه به طبقات مختلف ریسک مؤدیان انجام خواهد شد. افزایش رضایت صاحبان منافع از طریق برقراری عدالت در دریافت مالیات و تکریم مؤدیان مالیاتی از اهداف مهم این طرح است هزینه‌های بالای تمکین یعنی هزینه‌های که مؤدی باید برای گردآوری اطلاعات، تنظیم فرم‌های مالیاتی و غیره متحمل شود می‌تواند دلیل دیگری برای اجتناب و فرار مالیاتی باشد. شاخص توسعه جهانی^۱ بانک جهانی (۲۰۰۸) برای "زمان تنظیم و پرداخت مالیات" تفاوت فاحشی را در میان کشورها نمایش می‌دهد. این زمان بطور متوسط در کشورهای توسعه یافته و پر درآمد ۲۱۰ ساعت است در حالیکه در کشورهای مثل بولیوی و ویتنام این زمان به ۱۰۸۰ ساعت و در برزیل به ۲۶۰۰ ساعت می‌رسد. این وضعیت کسب و کارها را از بار اداری بیش از بار مالیاتی نگران می‌کند. عمده‌تاً کسب و کارهای کوچک و متوسط از هزینه‌های تمکین بالا بیشتر آسیب می‌بینند. (Iranian National Tax Administrations Research Database, 1401).

به سبب اختلاف بین قوانین و مقررات مالیاتی و اصول پذیرفته شده حسابداری ممکن است آنچه برای مقاصد حسابداری درآمد و هزینه محسوب می‌شود، الزاماً از دیدگاه قوانین مالیاتی درآمد و هزینه محسوب نگردد. علاوه بر این، برخی از اقلام درآمد و هزینه که از دیدگاه مقاصد حسابداری در یک دوره مالی شناسایی می‌شوند، ممکن است لحاظ قوانین مالیاتی در دوره دیگری شناسایی شوند. هرکدام از موارد فوق موجب بروز اختلاف بین سود حسابداری و درآمد مشمول مالیات شرکت می‌شود (Ghanbari fard, 1995). مبانی نظری و شواهد تجربی نشان می‌دهند شرکت‌ها سعی در کاهش و به تعویق انداختن مالیات بر درآمد خود دارند. اجتناب مالیاتی، فرار مالیاتی، مدیریت سود، محافظه‌کاری و غیره ابزارهایی هستند که شرکت‌ها ممکن است در صورت لزوم بسته به شرایط از آن‌ها

1. World Development Indicator

استفاده نمایند. شرکت‌ها انگیزه دارند تا با اعمال حسابداری محافظه‌کارانه در هزینه مالیات خود صرفه‌جویی کنند (Watts, 2003). حسابرسی مبتنی بر ریسک یکی از ماژول‌های طرح جامع مالیاتی است در این ماژول مؤدیان مالیاتی براساس میزان ریسک رتبه‌بندی می‌شوند و متناسب با سطح ریسک مؤدی (بالا، متوسط یا پایین) مؤدیان با ریسک بالا و نمونه‌ای از مؤدیان با ریسک متوسط حسابرسی مالیاتی می‌شوند و مالیات آنها از طریق برگ تشخیص ابلاغ می‌گردد و برای سایر مؤدیان مالیاتی بدون حسابرسی به صورت سیستمی برگه قطعی صادر می‌گردد. موتورهای ریسک براساس مؤلفه‌های گوناگونی مانند سوابق مالیاتی، اندازه شرکت، نسبت‌های مالی، مدیریت سود، جریان‌های وجه نقد، محافظه‌کاری، حاکمیت شرکتی و... رتبه اعتباری مؤدیان مالیاتی را تعیین می‌کنند.

باقری و باباجانی سال ۱۳۹۸ در تحقیقی با عنوان شناسایی عوامل مؤثر بر تمکین مالیاتی در ایران پس از انجام مطالعات اکتشافی و بررسی تحقیق‌های انجام شده و نیز با در نظر گرفتن شرایط محیطی حاکم بر کشور ایران، عوامل مذکور را در ۸ مؤلفه و ۴۰ شاخص طبقه‌بندی و با به کارگیری روش دلفی فازی، نظرات ۶۱ خبره درخصوص تأثیر هر یک از شاخص‌های مذکور بر تمکین مالیاتی را دریافت و مورد تحلیل قرار داده‌اند. نتایج تحقیق بیانگر آن بوده که از نظر خبرگان، کم بودن هزینه تمکین برای مؤدیان، سیستم کارآمد اطلاعات مالیاتی، رسیدگی و حسابرسی مالیاتی، پیچیدگی قوانین و مقررات بر تمکین مالیاتی اثرگذار می‌باشند.

محمد مسیحی و همکاران در سال ۱۳۹۸ در تحقیقی با عنوان استفاده از تکنیک‌های داده‌کاوی برای سنجش ریسک مالیاتی مؤدیان مالیات بر ارزش افزوده دریافتند که دو تکنیک ماشین بردار پشتیبان خطی و رگرسیون لجستیک از توان صحت ارزیابی ۷۰٪ برخوردار هستند و نوعی ادغام در نتایج این دو تکنیک توانسته است با کسب نزدیک به ۸۳٪ صحت ارزیابی از توان بالاتری برخوردار باشد نتایج حاصل از تحقیقات صالحی، صراف و مهرانی در سال ۱۴۰۰ نشان داد که از بین عوامل مؤثر مورد بررسی بر ریسک حسابرسی مالیاتی مؤدی، ارتباط بین ارائه اظهارنامه مالیاتی و دوره رسیدگی معاملات مؤدی با ریسک حسابرسی معکوس است، اما ارتباط بین حجم معاملات مؤدی با ریسک حسابرسی مستقیم است و همچنین نتایج بیانگر عدم وجود رابطه بین عضویت شرکت مؤدی در بورس و نسبت‌های سودآوری مؤدیان (بازده سرمایه در گردش، بازده سهام، سود هر سهم) با ریسک حسابرسی مالیاتی مؤدی بوده است. نیکو مرام، بنی مهد و احمدی لویه سال ۱۴۰۰ در تحقیق خود روی نمونه‌ای متشکل از ۱۷۳ شرکت فعال در بورس اوراق بهادار تهران در دوره زمانی ۱۳۸۷ الی ۱۳۹۶ به این نتیجه رسیده‌اند که شرکت‌های دارای حق انتخاب حسابرسی در انتخاب حسابرسی به منافع خود از جمله اجتناب مالیاتی توجه می‌کنند و حسابرسان نیز به دلیل اینکه با مسئله انتخاب مجدد در حسابرسی‌های بعدی همراه هستند، کوشش می‌کنند تا از طریق همکاری با این شرکت‌ها در زمینه اجتناب مالیاتی رضایت آنها را جلب نمایند و در نهایت این

جریان موجب می‌شود تا اجتناب مالیاتی برای شرکت‌های دارای حق انتخاب حسابرس، بیشتر از شرکت‌های فاقد حق انتخاب حسابرس باشد. نتایج مطالعه‌ای با عنوان جرایم مالیاتی و تمکین در شرکت‌های کوچک و متوسط مالزی، که توسط یونس، راملی و ابوحسن در سال ۲۰۱۷ انجام شده حاکی از این است که تمام عناصر جریمه‌های مالیاتی یعنی نرخ و مجازات با تمکین مالیاتی رابطه معناداری دارند.

داسیلوا و همکاران در سال ۲۰۱۹ به بررسی دو سیاست مالیات اجباری و مالیات داوطلبانه در چارچوب شیب لغزنده در کشور برزیل پرداختند. نتایج پژوهش آنها نشان می‌دهد تعامل مبتنی بر اعتماد بین مالیات دهندگان و مأموران مالیاتی منجر به پرداخت مالیات داوطلبانه خواهد شد. این در حالی است که، سیاست مبتنی بر فشار و زور منجر به پرداخت مالیات اجباری می‌شود و مالیات‌دهندگان تمایلی به پرداخت مالیات ندارند. فرانسیسکو جی دلگادو در سال ۲۰۲۳ در پژوهشی با عنوان «اجتناب مالیاتی و مدیریت سود: رویکرد شبکه عصبی برای اقتصادهای بزرگ اروپا» روابط بین اجتناب مالیاتی و مدیریت سود در پنج اقتصاد برتر اتحادیه اروپا را با استفاده از شبکه عصبی مصنوعی بررسی کرده‌اند در این پژوهش داده‌های کشورهای آلمان، بریتانیا، فرانسه، ایتالیا و اسپانیا برای دوره‌های سال ۲۰۰۶ الی ۲۰۱۵ با تمرکز بر اقالام تعهدی اختیاری تجزیه و تحلیل شد. در این پژوهش اجتناب مالیاتی با سه شاخص ارزیابی شده است (دو شاخص براساس نرخ مالیات مؤثر و یک شاخص براساس تفاوت درآمد دفتری و مالیاتی^۱) نتایج این پژوهش مشخص کرد که وقوع الگوهای غیرخطی مثبت و معنی دار بین اقالام تعهدی اختیاری و هر دو تعیین کننده نرخ مالیات مؤثر دلالت دارد بر اینکه وقتی شرکت‌ها برای مدیریت سود چاره‌جویی می‌کنند، درآمد مشمول مالیات بیشتر و بنابراین نرخ مؤثر مالیات بیشتر و اجتناب مالیاتی کمتر می‌شود. همچنین روشن شد که اقالام تعهدی اختیاری بر تفاوت درآمد دفتری و مالیاتی مؤثر نیست و مشاهدات ما پیشنهاد نمی‌کند که شرکت‌ها برای کاهش مالیات پرداختی از دستکاری مالیات بهره‌کشی می‌کنند. بنابراین به نظر می‌رسد مدیریت سود بر شکاف بین حسابداری و مالیات اثرگذار نیست.

فرضیه‌های پژوهش

فرضیه‌هایی که در این پژوهش برای دستیابی به مدل مناسب برای اعتبارسنجی مؤدیان مالیاتی مورد آزمون قرار گرفت، به شرح زیر است:

فرضیه اول: امکان پیش‌بینی ریسک مؤدیان مالیاتی با استفاده از مدل درخت تصمیم (الگوریتم C5.0 و جنگل تصادفی) وجود دارد.

فرضیه دوم: امکان پیش‌بینی ریسک مویدان مالیاتی با استفاده از الگوریتم SVM وجود دارد.

فرضیه سوم: روش درخت تصمیم برای اعتبارسنجی مؤدیان مالیاتی مناسب‌تر است

روش‌شناسی تحقیق

این پژوهش از نوع توصیفی و از نظر هدف کاربردی می‌باشد. جامعه آماری این پژوهش کل شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌باشد که از عملکرد سال ۱۳۹۲ لغایت ۱۳۹۷ در بورس فعال بوده‌اند. در این پژوهش برای انتخاب نمونه از روش غربالگری (حذفی) استفاده شده است به این ترتیب که شرکت‌های دارای شرایط زیر به عنوان نمونه انتخاب شده‌اند:

۱. حداقل از ابتدای سال ۱۳۹۲ در بورس اوراق بهادار تهران پذیرفته شده بود
۲. اطلاعات مورد نیاز جهت اندازه‌گیری متغیرها تحقیق در مورد آن شرکتها، در طول دوره تحقیق در دسترس بود.

در این پژوهش ابتدا با استفاده از روش دلفی و متاستز به شناسایی متغیرهای مؤثر بر ریسک مؤدیان مالیاتی پرداخته شده است، پس از گردآوری و جمع‌بندی نتایج با حذف موارد تکراری و تجمیع موارد مشابه، در نهایت ۱۶۴ مؤلفه در اعتبارسنجی مؤدیان مالیاتی، مؤثر شناخته شد. سپس داده‌های مربوط به متغیرهای طی سال‌های ۱۳۹۲-۱۳۹۷ از طریق سایت کدال و بررسی پرونده‌های مالیاتی مؤدیان گردآوری شده است. تعیین رتبه اعتباری مؤدیان براساس سوابق مالیاتی (درصد اختلاف بین مالیات قطعی شده و ابرازی) مشخص شده است در نهایت با استفاده از نرم‌افزار Matlab و SPSS Modeler دقت درخت تصمیم (الگوریتم C5.0، جنگل تصادفی) و ماشین بردار پشتیبان در محاسبه میزان ریسک مؤدی مورد آزمون قرار گرفت و از روش ROC، (مشخصه عملیاتی گیرنده) برای مقایسه صحت پیش‌بینی مدل‌ها استفاده شده است که در ادامه مراحل انجام کار به صورت مختصر توضیح داده می‌شود.

تعریف نظری و عملیاتی متغیرها

متغیر وابسته

اصطلاح ریسک به عنوان پراکندگی تعریف می‌شود که در میزان عدم اطمینان نسبت به آینده نمود پیدا می‌کند (Brealey & Myers, 1991). نیومن و همکاران مفهوم کلی ریسک را به ریسک مالیاتی شرکت مرتبط ساخته و بیان می‌کنند: ریسک مالیاتی، به پتانسیلی که یک فعالیت یا عمل، باعث ایجاد تفاوت بین مبلغ مالیات قطعی با مبلغ ابرازی می‌شود، اشاره دارد. تفاوت مالیات قطعی با ابرازی شرکت‌ها به واسطه عدم اطمینان نسبت به پرداخت مالیات در آینده، به عنوان ریسک مالیاتی در نظر گرفته شده است (Neuman et al, 2013, 9). از دیدگاه سازمان مالیاتی درصد انحراف مبلغ مالیات ابرازی طبق اظهارنامه از مبلغ مالیات قطعی شده ریسک اعتباری

تعریف می‌شود که این انحراف در این پژوهش به عنوان متغیر وابسته در نظر گرفته شده است. بنابراین این شکاف بین مبلغ مالیات ابرازی و قطعی هم می‌تواند ناشی از فرار باشد هم اجتناب که هر دو مورد در متغیر وابسته یعنی ریسک مالیاتی پوشش داده می‌شود.

متغیرهای مستقل

متغیرهای مستقل شناسایی شده بالغ بر ۱۶۴ مؤلفه هستند. در جدول شماره ۱ به برخی از متغیرها اشاره می‌شود.

جدول (۱) - متغیرهای مستقل

متغیر مستقل	روش محاسبه	توضیحات
نوع مبانی تعدیل گزارش حسابرسی سال قبل	مقبول ۱، مشروط ۲، عدم اظهارنظر ۳، مردود ۴	درجه اعتبار صورت‌های مالی مؤدی در سال گذشته
نرخ رشد مالیات سال قبل نسبت به دو سال قبل	$\frac{Tax_n - Tax_{n-1}}{Tax_{n-1}}$	درصد رشد مالیات مؤدی در سال‌های گذشته
اندازه شرکت	لگاریتم طبیعی دارایی‌ها	انتظار می‌رود مؤدیان با اندازه بزرگتر ریسک پایینتری داشته باشند
جریمه کتمان درآمد	به استناد ماده ۱۹۲ ق. م. م برابر است با ۳۰ درصد مالیات درآمد کتمان شده	درآمد کتمان شده درآمدی است که به موجب اسناد و مدارک تحقق آن قابل اثبات است لکن در اظهارنامه ابراز نشده است.
نسبت جریان وجه نقد عملیاتی به درآمد مشمول مالیات	حاصل تقسیم وجه نقد حاصل از فعالیت‌های عملیاتی بر درآمد مشمول مالیات	چگونگی ارتباط بین سودآوری واحد تجاری و توان آن جهت ایجاد وجه نقد
جرائم مؤدی	جمع جرائم مؤدی موضوع ماده ۱۹۰، ۱۹۲ و ۱۹۳ ق. م. م	بیان کننده میزان تمکین مؤدی به تکالیف مقرر شده بر اساس ق. م. م
محافظه کاری	$CONSV_{it} = -1 * \left(\frac{1}{3} \sum_{t=1}^3 \frac{NOACC_{it}}{TA_{it}} \right)$	مدل ارائه شده توسط گیولی و هاین (۲۰۰۰)
اقلام تعهدی اختیاری	DA= Accrual-NDA	اقلام تعهدی است که مدیریت می‌تواند کنترل‌هایی روی آنها اعمال کند

منبع: یافته‌های پژوهش

روش تحلیل داده‌ها و آزمون فرضیه

آزمون تحلیل واریانس

به منظور شناسایی متغیرهایی که ممکن بود در تعیین رتبه ریسک شرکت‌ها تأثیر داشته باشند از آزمون تحلیل واریانس استفاده شد. به این صورت که از بین شاخص‌های احصا شده در گام قبل، شاخص‌هایی که در هر سه گروه شرکت‌های با ریسک کم، ریسک متوسط و ریسک بالا اختلاف معنی‌داری داشته‌اند یا مؤثر بوده‌اند انتخاب و سایر متغیرها به دلیل داشتن میانگین یکسان در سطوح ریسک مختلف از ورود به مدل‌های منتخب کنار گذاشته شدند، در مجموع ۴۸ شاخص مؤثر انتخاب شد و در ادامه با استفاده از آنها ریسک اعتباری مؤدیان از طریق درخت تصمیم (الگوریتم C5.0 و جنگل تصادفی) و ماشین بردار پشتیبان برآورد می‌شود. به منظور آزمون مدل‌ها، ابتدا اطلاعات مورد نیاز برای محاسبه متغیرهای ورودی از سایت کدال و پرونده‌های مالیاتی استخراج گردیده است.

درخت تصمیم‌گیری^۱

درخت تصمیم‌گیری یک ابزار برای پشتیبانی از تصمیم است که از درخت‌ها (بر اساس قاعده اگر - آنگاه) برای مدل کردن داده‌ها استفاده می‌کند. درخت تصمیم به‌طور معمول در مطالعات و طرح‌های تحقیقاتی مختلفی به کار می‌رود. به‌طور خاص درخت تصمیم براساس قواعد تصمیم‌گیری برای پیش‌بینی و دسته‌بندی مورد استفاده قرار می‌گیرد. رویکرد درخت تصمیم که معمولاً به قیاس قواعد شناخته می‌شود مزایای متعددی دارد. از جمله مزایای آن این است که پس از ساخت درخت به راحتی می‌توان علت استنتاج قواعد به دست آمده را مشاهده کرد؛ به این معنی که درخت تصمیم همانند شبکه عصبی مانند جعبه سیاه عمل نمی‌کند و منطق کاری آن واضح و روشن است و از دیگر مزایای آن شناخت بهتر متغیرها یا ویژگی‌های با اهمیت است. در ادامه از روش‌های مختلف درخت تصمیم برای شناسایی متغیرهای مؤثر در تعیین وضعیت ریسک شرکت‌ها استفاده شد.

الگوریتم C5.0

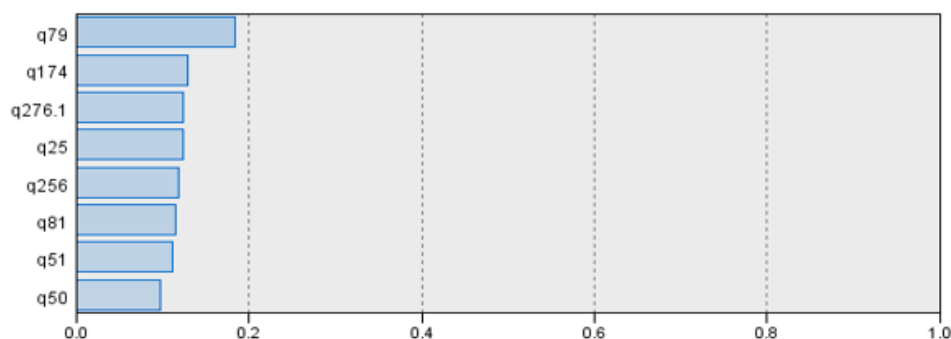
الگوریتم C5.0 با ۴ قاعده براساس دقت‌های مختلف محاسبه و قواعد آن بر اساس نحوه ترکیب^۲ ۶۰ درصد استخراج شد. با توجه به قاعده اول با دقت ۸۱.۴۱ درصد پیش‌بینی مشاهده می‌شود شرکت‌هایی که متغیر جریمه کتمان درآمد آنها مساوی صفر است (حدود ۱۳۷ شرکت)، اغلب کم ریسک هستند. از بین این شرکت‌ها حدود ۲۷ شرکت، شرکت‌هایی هستند که نسبت سود و زیان انباشته به کل دارای آنها کمتر مساوی (۰.۰۲۳-) است و اغلب پریسک هستند. از بین این شرکت‌ها ۲۴ شرکت مقدار نسبت خالص جریان وجه نقد عملیاتی به درآمد مشمول مالیات آنها کمتر مساوی ۲۰.۵۳۱ است که با اطمینان ۹۱.۷ درصد پریسک هستند. سه

1. Decision Tree

2. Boosting

شرکت دارای مقدار نسبت خالص جریان وجه نقد عملیاتی به درآمد مشمول مالیات بیشتر از ۲.۵۳۱ هستند و با اطمینان ۰.۶۶ درصد دارای ریسک متوسط هستند. شرکت‌هایی که نسبت سود انباشته به کل دارایی آنها بیش از (۰.۰۲۳-) است، اغلب آنها کم ریسک هستند (حدود ۱۱۰ شرکت). از این میان شرکت‌هایی که نسبت خالص جریان وجه نقد عملیاتی به درآمد مشمول مالیات آنها کمتر مساوی ۰.۷۸۸ بود اغلب دارای ریسک متوسط (۵۰ شرکت) هستند.

نمودار (۱) - میزان اهمیت هر یک از متغیرها با استفاده از الگوریتم C5.0



منبع: یافته‌های پژوهش

با توجه به نمودار بالا مشاهده می‌شود مهمترین متغیرها برای پیش‌بینی درخت تصمیم در تعیین ریسک شرکت‌ها به ترتیب جریمه کتمان درآمد، بازده دارایی‌ها (ROA)، نسبت جریان وجه نقد عملیاتی به درآمد مشمول مالیات، نسبت جمع بدهی‌ها به جمع دارایی‌ها، نسبت سود انباشته به کل دارایی، مالیات ابرازی سال $n-1$ ، مالیات سال $n-4$ و مالیات سال $n-3$ است.

جدول (۲) - جدول دوی بعدی متغیر وابسته و مقدار پیشگویی شده آن توسط الگوریتم درخت تصمیم C5.0

رتبه اعتباری	نتایج پیشگویی متغیر وابسته			
	ریسک بالا	ریسک متوسط	ریسک پایین	کل
ریسک پایین	تعداد	۱	۴۶	۵۰
	درصد سطر	۲	۹۲	۱۰۰
ریسک متوسط	تعداد	۶	۴	۴۳
	درصد سطر	۱۳.۹۵۳	۹.۳۰۲	۱۰۰
ریسک بالا	تعداد	۶۱	۰	۶۳
	درصد سطر	۸۲۵.۹۶	۰	۱۰۰
کل	تعداد	۷۰	۳۶	۱۵۶
	درصد سطر	۴۴.۸۷۲	۲۳.۰۷۷	۳۲.۰۵۱

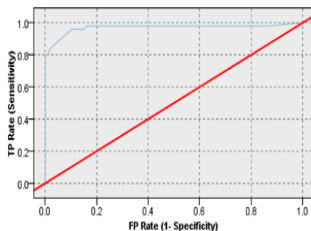
کای دو = ۱۹۹.۲۲۱ درجه آزادی = ۰.۴ احتمال = ۰

منبع: یافته‌های پژوهش

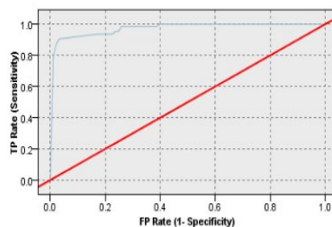
با توجه به جدول دوی بعدی حاصل از ریسک شرکت‌ها و ریسک شناسایی شده حاصل از پیش‌بینی توسط درخت تصمیم، دقت درخت تصمیم در این حالت حدود ۹۰ درصد و میزان خطای آن حدود ۱۰ درصد می‌باشد.

نمودار ROC^۱ برای الگوریتم C5.0

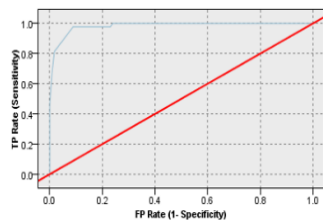
منحنی ROC یکی از مفیدترین و کارآمدترین روش‌ها در ارائه خصوصیت تعیینی، شناسایی احتمالی و پیش‌بینی مدل‌ها است که میزان دقت مدل را به صورت کمی برآورد می‌کند. سطح زیر منحنی ROC با عنوان AUC بیانگر مقدار پیش‌بینی مدل از طریق توصیف توانایی آن در تخمین درست وضعیت شرکت‌ها است.



نمودار ۴. منحنی ROC برای الگوریتم C5.0 (ریسک بالا)



نمودار ۳. منحنی ROC برای الگوریتم C5.0 (ریسک متوسط)



نمودار ۲. منحنی ROC برای الگوریتم C5.0 (ریسک پایین)

با استفاده از الگوریتم C5.0 مقدار AUC در شرکت‌های باریسک کم ۰.۹۷، در شرکت‌های با ریسک متوسط ۰.۹۸ و در شرکت‌های باریسک بالا ۰.۹۶ است. ایده‌آل‌ترین مدل، بیشترین سطح زیر منحنی را دارد و مقادیر AUC از صفر تا ۱ متغیر است. هر چه این مقدار به یک نزدیکتر باشد مدل مناسب‌تر است. با توجه به مقدار AUC (جدول شماره ۳) الگوریتم C5.0 به ترتیب در تشخیص شرکت‌های با ریسک متوسط، کم ریسک و پرریسک قابلیت بیشتری دارد.

جدول (۳) - محدوده زیر منحنی لگوریتم C5.0 *

متغیر	سطح ریسک	اندازه محدوده
	ریسک کم	٪۹۷
ریسک شرکت	ریسک متوسط	٪۹۸
	ریسک بالا	٪۹۶

منبع: یافته‌های پژوهش

الگوریتم جنگل تصادفی

جدول (۴) - اطلاعات مدل

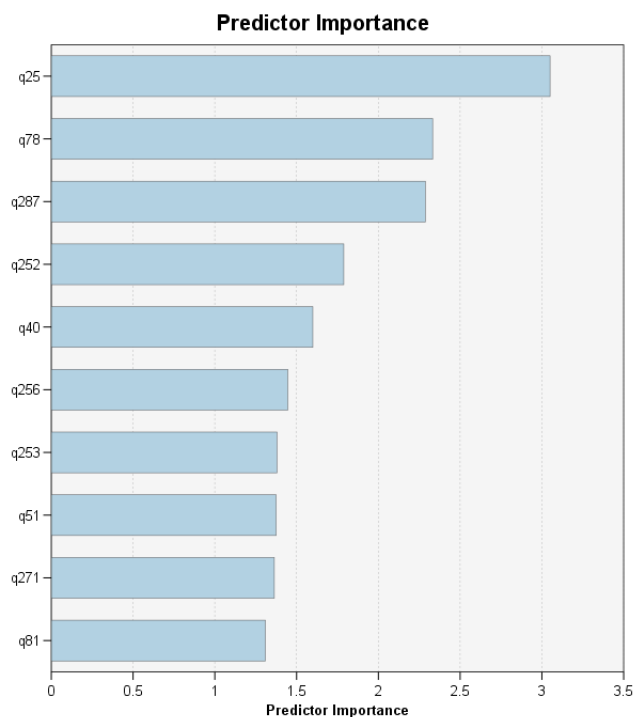
رتبه اعتباری	هدف
جنگل تصادفی	روش ساخت مدل
۴۱	تعداد ورودی‌های پیش‌بینی کننده
۰.۴۸۴	میانگین هندسی
۰.۵۲۰	۱
۰.۴۲۹	۲ نرخ موارد مثبت صحیح ^۱
۰.۵۰۸	۳

منبع: یافته‌های پژوهش

1. True Positive Rate

با توجه به آماره‌های ارائه شده توسط الگوریتم جنگل تصادفی مشاهده می‌شود نرخ تشخیص مشاهده‌های درست برای هر یک از ریسک‌ها به ترتیب از ریسک پایین تا ریسک بالا ۰.۴۲۹، ۰.۵۲۰، و ۰.۵۰۸ است. میانگین نرخ درست تشخیص ۰.۴۸۴ است.

نمودار (۵) - میزان اهمیت متغیرها در رشد درخت تصمیم با به کارگیری الگوریتم جنگل تصادفی



منبع: یافته‌های پژوهش

نمودارهای بالا به ترتیب متغیرهای مؤثر در رشد درخت، در الگوریتم درخت تصمیم را نشان می‌دهند، مهمترین متغیرها به ترتیب نسبت جمع بدهی‌ها به جمع دارایی‌ها، میزان جرایم مؤدی، نسبت سرمایه در گردش به کل دارایی‌ها، نسبت سود عملیاتی به فروش خالص، رشد موزون درآمد و بدهی (نسبت درآمد به بدهی)، نسبت سود انباشته به کل دارایی، نسبت سود و زیان قبل از بهره و مالیات به دارایی‌ها، مالیات سال $n-4$ ، نسبت سود عملیاتی به مجموع دارایی‌ها در سال $n-1$ و مالیات ابرازی سال $n-1$ است.

جدول (۵) - دو بعدی ریسک شرکتهای و نتایج حاصل از پیشگویی الگوریتم جنگل تصادفی

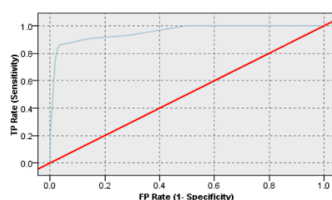
رتبه اعتباری	نتایج پیشگویی متغیر وابسته				کل
	ریسک پایین	ریسک متوسط	ریسک بالا	ریسک کل	
ریسک پایین	تعداد	۴۶	۳	۱	۵۰
	درصد سطر	۹۲	۶	۲	۱۰۰
ریسک متوسط	تعداد	۵	۳۶	۲	۴۳
	درصد سطر	۱۱.۶۲۸	۸۳.۷۲۱	۴.۶۵۱	۱۰۰
ریسک بالا	تعداد	۶	۱۱	۴۶	۶۳
	درصد سطر	۹.۵۲۴	۱۷.۴۶۰	۷۳.۰۱۶	۱۰۰
کل	تعداد	۵۷	۵۰	۴۹	۱۵۶
	درصد سطر	۳۶.۵۳۸	۳۲.۰۵۱	۳۱.۴۱۰	۱۰۰

کای دو = ۱۹۹.۲۲۱ درجه آزادی = ۴، احتمال = ۰

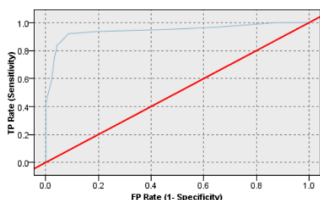
منبع: یافته‌های پژوهش

میزان دقت این الگوریتم ۸۲ درصد و میزان خطای آن ۱۸ درصد به دست آمد.

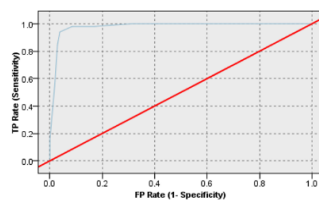
نمودار ROC برای الگوریتم جنگل تصادفی



نمودار ۸. منحنی ROC برای الگوریتم جنگل تصادفی (ریسک بالا)



نمودار ۷. ROC برای الگوریتم جنگل تصادفی (ریسک متوسط)



نمودار ۶. منحنی ROC برای الگوریتم جنگل تصادفی (ریسک پایین)

با استفاده از الگوریتم جنگل تصادفی، مقدار AUC در شرکتهای باریسک پایین ۰.۹۸، شرکتهای باریسک متوسط ۰.۹۸ است در شرکتهای باریسک بالا ۰.۹۳ است.

جدول (۶) - محدوده زیر منحنی

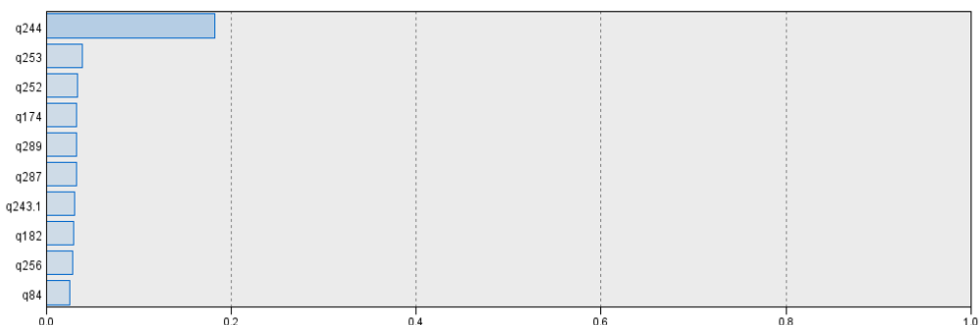
متغیر	سطح ریسک	اندازه محدوده
	ریسک کم	٪۹۸
ریسک شرکت	ریسک متوسط	٪۹۸
	پر ریسک	٪۹۳

منبع: یافته‌های پژوهش

با توجه به مقدار AUC (جدول شماره ۶) الگوریتم جنگل تصادفی به ترتیب در تشخیص شرکت‌های با ریسک متوسط، کم ریسک و با ریسک بالا قابلیت بیشتری دارد.

الگوریتم SVM'

نمودار (۹) - میزان اهمیت متغیرها در رشد درخت تصمیم با به کارگیری الگوریتم SVM



منبع: یافته‌های پژوهش

براساس نتایج حاصل از الگوریتم SVM مشاهده می‌شود مهمترین متغیرها برای تشخیص ریسک شرکت‌ها به ترتیب متغیرهای نوع مبانی تعدیل گزارش حسابرسی سال قبل، نسبت سود و زیان قبل از بهره و مالیات به دارایی‌ها، نسبت سود عملیاتی به فروش خالص، بازده دارایی‌ها (ROA)، نسبت خالص سرمایه در گردش به دارایی‌ها، نسبت سرمایه در گردش به کل دارایی‌ها، نسبت خالص جریان وجه نقد عملیاتی به درآمد مشمول مالیات، محافظه کاری، نسبت سود انباشته به کل دارایی و مالیات ابرازی سال $n-4$ بودند.

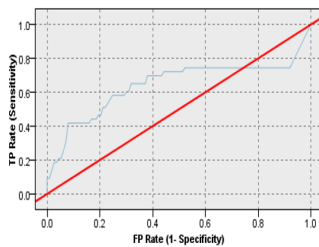
جدول (۷) - جدول دو بعدی ریسک شرکت‌ها و نتایج حاصل از پیشگویی الگوریتم SVM

رتبه اعتباری	نتایج پیشگویی متغیر وابسته				کل
	بی اثر ^۱	ریسک پایین	ریسک متوسط	ریسک بالا	
ریسک پایین	تعداد	۵	۳۱	۴	۵۰
	درصد سطر	۱۰	۶۲	۸	۱۰۰
ریسک متوسط	تعداد	۱۱	۴	۱۸	۴۳
	درصد سطر	۲۵.۵۸۱	۹.۳۰۲	۴۱.۸۶۰	۱۰۰
ریسک بالا	تعداد	۴	۸	۶	۶۳
	درصد سطر	۶.۳۴۹	۱۲.۶۹۸	۹.۵۲۴	۱۰۰
کل	تعداد	۲۰	۴۳	۲۸	۱۵۶
	درصد سطر	۱۲.۸۲۱	۲۷.۵۶۴	۱۷.۹۴۹	۱۰۰

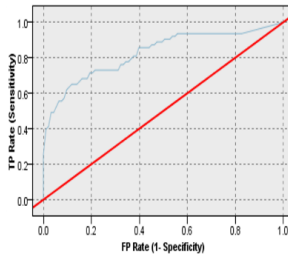
کای دو = ۰.۰۶۴، ۸۱.۰۶۴، درجه آزادی = ۶، احتمال = ۰

منبع: یافته‌های پژوهش

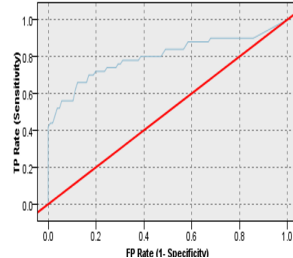
نمودار ROC برای الگوریتم SVM



نمودار ۱۲. منحنی ROC برای الگوریتم SVM (ریسک بالا)



نمودار ۱۱. منحنی ROC برای الگوریتم SVM (ریسک متوسط)



نمودار ۱۰. منحنی ROC برای الگوریتم SVM (ریسک پایین)

با استفاده از الگوریتم SVM، مقدار AUC در شرکت‌های باریسک کم ۰.۷۸، در شرکت‌های باریسک متوسط ۰.۶۷ و در شرکت‌های باریسک بالا ۰.۸۳ است.

جدول (۸) - محدوده زیر منحنی

متغیر	سطح ریسک	اندازه محدوده
	ریسک کم	٪۷۸
ریسک شرکت	ریسک متوسط	٪۶۷
	پر ریسک	٪۸۳

منبع: یافته‌های پژوهش

با توجه به مقدار AUC (جدول شماره ۸) الگوریتم SVM به ترتیب در تشخیص شرکت‌های با ریسک بالا، کم ریسک و ریسک متوسط قابلیت بیشتری دارد.

مقایسه الگوریتم‌ها در اعتبارسنجی مؤدیان مالیاتی

جدول (۹) - مقایسه روش‌ها مختلف در تشخیص ریسک شرکت‌ها

مدل/روش	میزان دقت	خطا	مهمترین متغیرها
			۱ ۲ ۳
الگوریتم C5.0	٪۰.۹۰	٪۱۰	نرخ بازده دارایی‌ها (ROA) نسبت خالص جریان وجه نقد عملیاتی به درآمد مشمول مالیات
الگوریتم جنگل تصادفی	٪۰.۸۲	٪۰.۱۸	نسبت جمع بدهی‌ها به جمع دارایی‌ها میزان جرایم مؤدی
ماشین بردار پشتیبان	٪۰.۶۹	٪۰.۳۰	نسبت سود و زیان قبل از بهره و مالیات به دارایی‌ها نوع مبنای تعدیل گزارش حسابرسی سال قبل

منبع: یافته‌های پژوهش

جدول ۹ نشان می‌دهد الگوریتم‌های C5.0، جنگل تصادفی و ماشین بردار پشتیبان به ترتیب بیشترین دقت را در اعتبارسنجی مؤدیان مالیاتی دارند.

مقایسه روش‌ها اعتبارسنجی مؤدیان مالیاتی بر اساس سطح زیر منحنی ROC

جدول زیر مقایسه همه الگوریتم‌ها بر اساس سطح زیر منحنی ROC را نشان می‌دهد.

جدول (۱۰) - مقایسه روش‌ها اعتبارسنجی مؤدیان مالیاتی بر اساس سطح زیر منحنی ROC

ردیف	الگوریتم	متغیر	سطح ریسک	اندازه محدوده	رتبه ارزیابی در سطح ریسک
۱	C5. 0	ریسک مالیاتی شرکت	پایین	٪۹۷	۲
			متوسط	٪۹۸	۱
			بالا	٪۹۶	۱
۲	جنگل تصادفی	ریسک مالیاتی شرکت	پایین	٪۹۸	۱
			متوسط	٪۹۸	۱
			بالا	٪۹۳	۲
۳	SVM	ریسک مالیاتی شرکت	پایین	٪۷۸	۳
			متوسط	٪۶۷	۳
			بالا	٪۸۳	۳

منبع: یافته‌های پژوهش

جدول فوق نشان می‌دهد که الگوریتم C5. 0 در شناسایی مؤدیان با ریسک بالا و متوسط بهترین عملکرد را دارد و در شناسایی مؤدیان با ریسک پایین رتبه دوم را دارد. همچنین مدل جنگل تصادفی در شناسایی مؤدیان با ریسک متوسط و پایین بهترین عملکرد را دارد و در شناسایی مؤدیان با ریسک بالا رتبه دوم را دارد و الگوریتم SVM در شناسایی مؤدیان در هر سه سطح ریسک دارای رتبه سوم می‌باشد.

جدول زیر رتبه‌بندی الگوریتم‌های مختلف شناسایی ریسک مؤدیان بر اساس سطح زیر منحنی ROC را در سه سطح ریسک بالا، متوسط و پایین نشان می‌دهند.

جدول (۱۱) - رتبه الگوریتم‌های مختلف در شناسایی مؤدیان با ریسک بالا

ردیف	الگوریتم	متغیر	سطح ریسک	اندازه محدوده	رتبه ارزیابی در سطح ریسک
۱	C5. 0	ریسک مالیاتی	بالا	٪۹۶	۱
۲	جنگل تصادفی	ریسک مالیاتی	بالا	٪۹۳	۲
۳	SVM	ریسک مالیاتی	بالا	٪۸۳	۳

منبع: یافته‌های پژوهش

جدول شماره ۱۱ نشان می‌دهد که الگوریتم C5.0 بهترین عملکرد را در شناسایی مؤدیان با ریسک بالا دارد سپس به ترتیب جنگل تصادفی و SVM بهترین عملکرد را در شناسایی مؤدیان با ریسک بالا دارند.

جدول (۱۲) - رتبه الگوریتم‌های مختلف در شناسایی مؤدیان با ریسک متوسط

ردیف	الگوریتم	متغیر	سطح ریسک	اندازه محدوده	رتبه ارزیابی در سطح ریسک
۱	C5.0	ریسک مالیاتی	متوسط	٪۹۸	۱
۲	جنگل تصادفی	ریسک مالیاتی	متوسط	٪۹۸	۲
۳	SVM	ریسک مالیاتی	متوسط	٪۶۷	۳

منبع: یافته‌های پژوهش

جدول شماره ۱۲ نشان می‌دهد الگوریتم C5.0 و جنگل تصادفی بهترین عملکرد را در شناسایی مؤدیان با ریسک متوسط دارند و SVM نیز عملکرد مناسبی در شناسایی مؤدیان با ریسک متوسط دارد.

جدول (۱۳) - رتبه الگوریتم‌های مختلف در شناسایی مؤدیان با ریسک پایین

ردیف	الگوریتم	متغیر	سطح ریسک	اندازه محدوده	رتبه ارزیابی در سطح ریسک
۱	جنگل تصادفی	ریسک مالیاتی	پایین	٪۹۸	۱
۲	C5.0	ریسک مالیاتی	پایین	٪۹۷	۲
۳	SVM	ریسک مالیاتی	پایین	٪۷۸	۳

منبع: یافته‌های پژوهش

جدول شماره ۱۳ نشان می‌دهد الگوریتم جنگل تصادفی بهترین عملکرد را در شناسایی مؤدیان با ریسک پایین دارد سپس به ترتیب C5.0 و SVM مناسبترین عملکرد را در شناسایی ریسک مؤدیان با ریسک پایین دارند.

یافته‌های پژوهش

با عنایت به تغییر قوانین مالیاتی و پیاده‌سازی طرح جامع مالیاتی، سنجش میزان ریسک مؤدیان مالیاتی و اجرای حسابرسی مبتنی بر ریسک از اهمیت ویژه‌ای برخوردار است. همانطور که پیشتر گفته شد مدل جامعی به منظور سنجش اعتبار مؤدیان طراحی و تبیین نشده است. برای حل این مسأله در این پژوهش ابتدا با استفاده از روش‌های دلفی و متاستز ۱۶۴ متغیر به عنوان مؤلفه‌های مؤثر بر اعتبارسنجی مؤدیان مالیاتی شناسایی شد و در گام بعد اطلاعات مورد نیاز برای آزمون مدل‌های درخت تصمیم و ماشین بردار پشتیبان از سایت کدال و همچنین مطالعه پرونده‌های مالیاتی مؤدیان مالیاتی استخراج شد. نتایج به دست آمده از آزمون داده‌ها با استفاده از الگوریتم C5.0 نشان داد دقت درخت تصمیم در این حالت حدود ۹۰ درصد و میزان خطای آن حدود ۱۰ درصد می‌باشد. با انجام این آزمون مشخص شد مهمترین متغیرها برای پیش‌بینی درخت تصمیم در تعیین ریسک شرکت‌ها به ترتیب جریمه کتمان درآمد، بازده دارایی‌ها (ROA)، نسبت خالص جریان وجه نقد عملیاتی به درآمد مشمول مالیات، نسبت جمع بدهی‌ها به جمع دارایی‌ها، نسبت سود انباشته به کل دارایی، مالیات ابرازی سال $n-1$ ، مالیات سال $n-4$ و مالیات سال $n-3$ است. نتایج به دست آمده از آزمون داده‌ها با استفاده از الگوریتم جنگل تصادفی نشان می‌دهد دقت درخت تصمیم در این حالت حدود ۸۲ درصد و میزان خطای آن حدود ۱۸ درصد می‌باشد و متغیرهای مؤثر در رشد درخت به ترتیب نسبت جمع بدهی‌ها به جمع دارایی‌ها، میزان جرایم مؤدی، نسبت سرمایه در گردش به کل دارایی‌ها، نسبت سود عملیاتی به فروش خالص، رشد موزون درآمد و بدهی (نسبت درآمد به بدهی)، نسبت سود انباشته به کل دارایی، نسبت سود و زیان قبل از بهره و مالیات به دارایی‌ها، مالیات سال $n-4$ ، نسبت سود عملیاتی به مجموع دارایی‌ها در سال $n-1$ و مالیات ابرازی سال $n-1$ هستند. نتایج به دست آمده از آزمون داده‌ها با استفاده از الگوریتم ماشین بردار پشتیبان نشان می‌دهد دقت این روش حدود ۶۹.۱ درصد و میزان خطای آن حدود ۳۰.۹ درصد می‌باشد و مهمترین متغیرها برای تشخیص ریسک شرکت‌ها به ترتیب متغیرهای نوع مبانی تعدیل گزارش حسابرسی سال قبل، نسبت سود و زیان قبل از بهره و مالیات به دارایی‌ها، نسبت سود عملیاتی به فروش خالص، بازده دارایی‌ها (ROA)، نسبت خالص سرمایه در گردش به دارایی‌ها، نسبت سرمایه در گردش به کل دارایی‌ها، محافظه‌کاری، نسبت سود انباشته به کل دارایی و مالیات ابرازی سال $n-4$ هستند.

بحث و نتیجه گیری

تحلیل نتایج نشان می‌دهد که الگوریتم‌ها عوامل متفاوتی را در اعتبارسنجی مؤدیان مالیاتی مؤثر تشخیص داده‌اند. این موضوع می‌تواند نشان‌دهنده این باشد که در شرایط مختلف ممکن است مؤلفه‌های متفاوتی در اعتبارسنجی مؤدیان مؤثر باشد. بنابراین هر اندازه تعداد مؤلفه‌ها بیشتر باشد و جنبه‌های مختلف رفتار مؤدیان ارزیابی شود، دقت اعتبارسنجی بیشتر خواهد بود. نرخ بازده دارایی‌ها (ROA) در مدل‌های CHAID، C5.0، الگوریتم SVM، الگوریتم شبکه عصبی جز مهمترین متغیرها بوده است. متغیر کنمان در سه مدل C5.0، CHAID، CART و SVM جز مهمترین متغیرها بوده است. نسبت خالص جریان وجه نقد عملیاتی به درآمد مشمول مالیات در مدل‌های C5.0 و الگوریتم SVM جز مهمترین متغیرها بوده است و نوع مبانی تعدیل گزارش حسابرسی سال قبل در الگوریتم‌های الگوریتم SVM و الگوریتم SVM جز مهمترین متغیرها است و نسبت جمع بدهی‌ها به جمع دارایی‌ها در مدل‌های C5.0 و الگوریتم جنگل تصادفی جز مهمترین متغیرها هستند. پرهزینه‌ترین نوع خطا در اعتبارسنجی مؤدیان مالیاتی تشخیص اشتباه مؤدیان با ریسک بالاست به این دلیل که در این صورت ممکن است مالیات مؤدیان با ریسک بالا بدون حسابرسی مالیاتی قطعی شود و احتمال فرار مالیاتی افزایش یابد. نتایج پژوهش نشان می‌دهد به ترتیب الگوریتم C5.0، جنگل تصادفی و SVM در تشخیص مؤدیان با ریسک بالا بهترین عملکرد را دارند.

همچنین هزینه خطا در شناسایی مؤدیان با ریسک پایین در صورت تشخیص به عنوان مؤدی با ریسک متوسط می‌تواند قرار گرفتن در لیست مؤدیان نمونه و در نهایت احتمال حسابرسی مالیاتی و تحمیل هزینه حسابرسی مالیاتی باشد و در صورت تشخیص به عنوان مؤدی پریسک نیز حسابرسی مالیاتی و در نهایت افزایش هزینه حسابرسی مالیاتی خواهد بود. بنابراین تشخیص اشتباه مؤدیان با ریسک پایین صرفاً باعث افزایش هزینه‌های حسابرسی مالیاتی می‌شود و موجب فرار مالیاتی نخواهد شد. نتایج نشان می‌دهد به ترتیب الگوریتم جنگل تصادفی، C5.0 و SVM در تشخیص مؤدیان با ریسک پایین بهترین عملکرد را دارند.

هزینه خطا در شناسایی مؤدیان مالیاتی با ریسک متوسط در صورت تشخیص به عنوان مؤدی با ریسک پایین، قرار نگرفتن مؤدی در فهرست مؤدیان نمونه و قطعی شدن پرونده بدون حسابرسی مالیات است که در نهایت افزایش احتمال فرار مالیاتی را به دنبال خواهد داشت و هزینه خطا در صورت تشخیص اشتباه مؤدی با ریسک متوسط به عنوان مؤدی پریسک نیز حتمی شدن حسابرسی مالیاتی مؤدی و افزایش هزینه‌های حسابرسی مالیاتی خواهد بود. نتایج پژوهش نشان می‌دهد به ترتیب الگوریتم C5.0، جنگل تصادفی و SVM در تشخیص مؤدیان با ریسک متوسط بهترین عملکرد را دارند. بنابراین با توجه به اینکه هدف اصلی سازمان مالیاتی جلوگیری از فرار مالیاتی می‌باشد، الگوریتم‌های که در شناسایی مؤدیان با ریسک بالا توانمندی بالاتری دارند برای اعتبارسنجی

مؤدیان مالیاتی مناسبتر هستند و با توجه به اینکه خطا در شناسایی مؤدیان با ریسک پایین، صرفاً باعث افزایش هزینه‌های رسیدگی می‌شود و اثری بر فرار مالیاتی ندارد بنابراین بهترین الگوریتم آن است که به ترتیب در شناسایی مؤدیان با ریسک بالا، متوسط و پایین توانمندی بیشتری دارد.

در مجموع بر اساس نتایج حاصل از مقدار AUC الگوریتم C5.0 و جنگل تصادفی برآزش بهترین دارند. با این وجود فرضیه پژوهش مبنی بر امکان پیش‌بینی ریسک مؤدیان مالیاتی با استفاده از الگوریتم SVM رد نمی‌شود.

در بین متغیرهای وارد شده در مدل نسبت جمع بدهی‌ها به جمع دارایی‌ها، جریمه کتمان درآمد، تفاوت بین مالیات ابرازی و تشخیصی مؤدی در سال $n-1$ ، تفاوت بین مالیات ابرازی و تشخیصی در سال $n-3$ ، تفاوت بین مالیات ابرازی و تشخیصی در سال $n-5$ ، نوع مبانی تعدیل گزارش حسابرسی سال قبل (مقبول ۱، مشروط ۲، عدم اظهارنظر ۳، مردود ۴)، تفاوت بین مالیات ابرازی و تشخیصی در سال $n-4$ ، نرخ رشد مالیات سال $n-1$ نسبت به سال $n-2$ ، تفاوت بین مالیات ابرازی و تشخیصی در سال n ، اندازه شرکت (لگاریتم طبیعی دارایی‌ها)، تفاوت بین مالیات ابرازی و تشخیصی در سال $n-2$ و... بیشترین اثر گذاری را در شناسایی ریسک مؤدیان مالیاتی دارند. همانطور که بحث شد پرهزینه‌ترین نوع خطا در اعتبارسنجی مؤدیان مالیاتی تشخیص اشتباه مؤدیان با ریسک بالاست به این دلیل که در این صورت ممکن است مالیات مؤدیان با ریسک بالا بدون حسابرسی مالیاتی قطعی شود و احتمال فرار مالیاتی افزایش یابد بنابراین خطا در شناسایی ریسک مؤدیان، به ترتیب در سطح ریسک بالا، متوسط و پایین بیشترین هزینه را برای سامان مالیاتی دارد.

بنابراین با توجه به موارد مذکور فرضیه‌های اول و دوم مبنی بر امکان پیش‌بینی ریسک مؤدیان مالیاتی با استفاده از مدل درخت تصمیم (الگوریتم C5.0، جنگل تصادفی) و ماشین بردار پشتیبان تأیید می‌گردند و در پاسخ به فرضیه سوم مبنی بر اینکه کدام یک از الگوریتم‌های درخت تصمیم (الگوریتم C5.0، جنگل تصادفی) و ماشین بردار پشتیبان برای اعتبارسنجی مؤدیان مالیاتی مناسب‌تر هستند با توجه به نتایج بدست آمده بهترین الگوریتم برای اعتبارسنجی مؤدیان مالیاتی به ترتیب الگوریتم C5.0، جنگل تصادفی و ماشین بردار پشتیبان هستند. نتایج حاصله با پژوهش‌های راویسانکار و همکاران در سال ۲۰۱۱؛ یونس، راملی و ابوحسن در سال ۲۰۱۷؛ محمد مسیحی و همکاران در سال ۱۳۹۸؛ ابراهار تینگر و همکاران (۲۰۲۰)، دستگیر و همکاران (۱۳۹۳)، مطابق است اما نتایج حاصله با پژوهش فرانسیسکو جی دلگادو در سال ۲۰۲۳ مطابقت ندارد. پیشنهادت زیر در راستای انجام پژوهش‌های آتی ارائه می‌شود:

۱. به روز رسانی مداوم شاخص‌های اندازه‌گیری ریسک با توجه به تغییرات در قوانین مالیاتی، استانداردهای حسابداری، رفتار مؤدیان و... و استفاده از شاخص‌هایی که کمتر مورد دستکاری واقع می‌شوند و ارتباط

- مستقیم با درآمد مشمول مالیات دارند مانند جریان وجه نقد عملیاتی
۲. با توجه به نتایج حاصل از فرضیه سوم پیشنهاد می‌شود سازمان امور مالیاتی از الگوریتم‌های درخت تصمیم مانند C5.0 و جنگل تصادفی برای اعتبارسنجی مؤدیان مالیاتی استفاده کند.
 ۳. پیشنهاد می‌شود سازمان امور مالیاتی برای تسهیل اعتبارسنجی، اظهارنامه‌های مالیاتی مؤدیان را به شکلی طراحی کند که اطلاعات مورد نیاز برای اعتبارسنجی مؤدیان مالیاتی در آن ثبت گردد.
 ۴. برای اعتبارسنجی سریع اظهارنامه، پیشگیری از آرایش حساب‌ها و افزایش سرعت رسیدگی‌های سازمان امور مالیاتی پیشنهاد می‌شود صورت‌مجلس تحویل اسناد و مدارک مؤدی نیز به صورت الکترونیکی به همراه اظهارنامه تا تاریخ انقضا مهلت تسلیم اظهارنامه انجام شود بدین ترتیب اطلاعات تکمیلی مورد نیاز برای اعتبارسنجی مؤدی مالیاتی از طریق صورت‌مجلس تا زمان انقضا مهلت تسلیم اظهارنامه در دسترس سازمان قرار خواهد گرفت و فرصت حساب‌سازی و... نیز کاهش خواهد یافت.
 ۵. پیشنهاد می‌شود برای دستیابی سریع و ساده به تمامی اطلاعات مورد نیاز برای اعتبارسنجی مؤدیان مالیاتی سازمان امور مالیاتی، سامانه یکپارچه‌ای طراحی کند که تمامی اطلاعات مربوط به گذشته و حال مؤدی (شامل سوابق پرونده، اطلاعات اظهارنامه مالیاتی، اطلاعات پرونده در تمامی مراجع دادرسی مالیاتی، اطلاعات اشخاص ثالث، اطلاعات سامانه‌های صورت معاملات فصلی، صورت‌های مالی حسابرسی شده سامانه جام)، اطلاعات مربوط به دارایی‌ها، و املاک و... در آن ثبت گردد.
 ۶. تأکید بر روی سوابق مالیاتی مؤدی از ابتدای اجرای طرح جامع مالیاتی به عنوان شاخص اعتبارسنجی مؤدیان مالیاتی و استفاده از جرائم مؤدی از قبیل جریمه کتمان درآمد و... به عنوان شاخص اعتبارسنجی مؤدیان.
 ۷. افزایش انجام کنترل‌های متقابل در حجم وسیع برای تأیید اطلاعات گزارش شده در اظهارنامه مالیاتی.
 ۸. تحلیل نتایج حاصل از کنکاش‌های محیطی که توسط سازمان امور مالیاتی به عنوان بخشی از برنامه‌ریزی استراتژیک برای شناسایی ریسک‌های نوظهور اتخاذ می‌شوند.
 ۹. شناسایی ریسک‌های تشخیص و رتبه‌بندی در چارچوب گروه‌بندی مؤدیان (یعنی در جایی که از اصول بخش‌بندی بازار برای تقسیم جمعیت مؤدیان در گروه‌های کوچک و قابل مدیریت مبتنی بر مشخصه‌های مشترک و ریسک‌ها استفاده می‌شود)، مالیات‌های اصلی و تکالیف مهم (ثبت نام، تنظیم اظهارنامه و گزارش‌دهی)
 ۱۰. داشتن سیستم ثبت ریسک (یعنی یک مخزن مرکزی از ریسک‌های شناسایی شده که تداوم عملیات سازمان امور مالیاتی را تهدید می‌کند). سیستم ثبت ریسک ممکن است از زمانی به سازمان دیگر متفاوت

باشد، اما عمدتاً به عنوان حداقل، اطلاعات زیر را در برخواهد داشت: شرح مختصر ریسک؛ تاریخ شناسایی ریسک؛ احتمال وقوع؛ شدت تاثیر آن؛ سنجه‌های کاهش ریسک؛ نام مالک ریسک (یعنی شخصی که پرداختن به ریسک را اطمینان می‌دهد)؛ و وضعیت ریسک.

سازمان امور مالیاتی می‌تواند از نتایج این پژوهش در مازول موتور ریسک طرح جامع مالیاتی برای توسعه شاخص‌ها و مدل اعتبارسنجی استفاده کند همچنین می‌تواند از پیشنهاد‌های این پژوهش برای محدود کردن گریزهای مالیاتی و افزایش کیفیت رسیدگی‌ها استفاده کند.

به منظور پیاده‌سازی سیستم اعتبارسنجی مؤدیان مالیاتی پیشنهاد می‌شود سایر الگوهای پیش‌بینی چون الگوریتم‌های فراابتکاری در بهینه‌سازی ترکیبی و مدل‌های محقق ساخته و الگوریتم ژنتیک و مورچگان مورد ارزیابی قرار بگیرد؛ تا بتوان دقت روش‌های مختلف را سنجید و روشی مناسب با کمترین خطای ممکن تبیین نمود. همچنین شناسایی ریسک‌های تشخیص و رتبه‌بندی در چارچوب گروه‌بندی‌های مؤدیان مالیاتی در ایران می‌تواند به دقت بیشتر رتبه‌بندی ریسک مؤدیان مالیاتی منجر شود.

فهرست منابع

1. Alabede, James O. , Zaimah Zainol Ariffin and Kamil Md Idris. (2011). Individual Taxpayers Attitude and Compliance Behaviour in Nigeria: The Moderating Role of Financial Condition and Risk Preference, Journal of Accounting and Taxation Vol. 3(5), pp. 91- 104.
2. Bagherpour M. A. Bagheri M. Khadem H. Hosieni Pour R. (2012). Examine the Effects of Financial and Non-Financial Variables on Tax Evasion Using Data Mining Techniques: Automotive and Parts Manufacturing Industry. Empirical Studies in Financial Accounting Quarterly, 34: 103- 12, (In Persian)
3. Brealey, R. and s. Myers (1991). Principles of Corporate Finance, Furth Edition, New York, McGraw-Hill, Inc.
4. Dastgir M, Izadinia N, Askari A, Ramezani M M. (2015). Providing a Model for Corporate Risk- based Audit Selection in Iran. Journal of Tax Research. 23 (25), (In Persian).
5. Dastgir M, Qaribi M. (2016). Using Data Mining Techniques to Enhance Tax Evasion Detection Performance. Journal of Tax Research, 23 (28), (In Persian).
6. Francisco J. Delgado, Elena Fernández-Rodríguez, Roberto García-Fernández, Manuel Landajo, Antonio Martínez-Arias (2023). Tax Avoidance and Earnings Management: a Neural Network Approach for the Largest European Economies, Financial Innovation volume 9, Article no. 19.
7. Gallemore, J. , and E. labro. (2015). The Importance of the Internal Information Environment for Tax Avoidance. Journal of Accounting and Economics. 60(1): 149-167.
8. Hosieni Pour R. Bagherpour M. Lashani M. A. Salehi M. (2017). Identify Financial and Non-Financial Variables Affecting the Basis of Auditing Report Adjustment Related to Accounting Estimates: A Data Mining Approach. Audit Science. 17(66): 107-130, (In Persian).
9. Iranian National Tax Administrations Research Database (2022). Approaches to Deal with Tax Evasion and Avoidance in Selected Countries, (In Persian).
10. Khoramnia H. , khoramnia A A, Mehrkam M. (2017). The Influence of Conservatism on Relationship between Operational Cash Flow and Definitive Taxable Income. Journal of Tax Research. 25 (35): 128-156, (In Persian).
11. Masihi M, Yaghobnejad A, Kaighobadi A, Torabi T, (2019). Using Data Mining

- Techniques to Measure Tax Risk, Value Added Taxpayers, Iranian Financial Engineering Association, (In Persian).
12. Namazi M, Sadeghzadeh Maharluie M. (2018). Predicting Tax Evasion by Decision Tree Algorithms. Quarterly Financial Accounting Journal. 9 (36): 76-101, (In Persian)
 13. Neuman, S. , T. Omer and A. Schmidt (2013). Risk and Return: Dose Tax Risk Reduce Firms Effective Tax Rates?, Working Paper, Texas A&M University and North Carolina State University.
 14. OECD, (2004). Compliance Risk Management: Audit Case Selection System, 2004, (In Persian).
 15. Pourzamani Z, Shamsi Jamkhaneh A. (2009). Factors Involved in the Difference between Taxable Income Declared by Business Corporations and Taxable Income Assessed by Tax Authorities: Case of West Tehran Tax Affairs Head Office. Journal of Tax Research. 17 (5): 9-26, (In Persian).
 16. Rahimikia E, Mohamadi S, Ghazanfari M. (2015). Tax Evasion Detection by Using Combinatory Elligent System. Journal of Tax Research. 23 (26): 136-164. (In Persian).
 17. Ramezani S M, Dastgir M, Asgari A, khani A. (2013). Risk-Based Tax Audit and its Determinants by Emphasizing on Taxpayers Size and other Countries Experience . J Tax Res. 21 (18) :77-114
 18. Taghavi Fard M, Raeesi Vanani I, Panahi R. (2017). A Predictive Analytics for Detection of VAT Taxpayers Evasion through Classification & Clustering Algorithms. Journal of Tax Research. 25 (35): 11-36, (In Persian).
 19. Watts, R. (2003). Conservatism in Accounting Part I: Explanations and Implications. Accounting Horizons 17(3): 207-222.
 20. Kirchler, E. (2007). The Economic Psychology of Tax Behavior. Cambridge: Cambridge University Press.
 21. De Ghanbari nia. Mihir A. , Alexander Dyckb, Luigi Zingalesc, ch. (2007). Theft and Taxes. Journal of Financial Economics 84591–623, (In Persian).