ارائه الگوی پیش‌بینی فرار مالیاتی اصناف مبتنی بر تکنیک داده‌کاوی

محمد قاسمی[۱]
صادق عابدی[۲]
علی محتشمی[۳]

چکیده

در این پژوهش با توجه به اهمیت موضوع و خلأ پژوهش‌های پیشین، یک مدل پیش‌بینی فرار مالیاتی اصناف مبتنی بر تکنیک داده کاوی ارائه می‌گردد. داده‌های مورد تحلیل شامل بررسی ۵۶۰۰ پرونده مالیاتی کلیه اصناف دارای کد مالیاتی در استان قزوین طی سال‌های ۹۸-۱۳۹۳ می‌باشد. پرونده مالیاتی مرتبط با اصناف در پنج گروه مالیاتی شامل گروه صنفی صاحبان دفاتر رسمی، گروه صنفی مشاورین املاک، گروه صنفی تالارهای پذیرایی، رستوان و مشاغل وابسته، گروه صنفی خدمات ارتباطی و گروه صنفی نمایشگاه و فروشگاه لوازم خودرویی و مشاغل وابسته می‌باشند. جهت مدل‌سازی از الگوی کلاس‌بندی درخت تصمیم استفاده گردید. نتایج نشان می‌دهد، مدل درخت تصمیم بر اساس داده‌های موجود، مدل مناسبی جهت پیش‌بینی محسوب می‌شود. معیار پوشش برابر ۶۸ ٪، معیار کاپا برابر ۰٫۶۱۲ بدست آمده است که عملکرد خوب مدل ساز را نشان می‌دهد. همچنین با استفاده از تکنیک Cross Validation صحت اعتبار مدل پیش‌بینی مورد آزمون قرار گرفت تا با اطمینان بیشتری درصد عملکرد مدل‌سازی تخمین زده شود. معیار صحت برابر ۶۷٫۷۹٪ نشان از قابلیت اطمینان مناسب جهت مدل پیش‌بینی می‌باشد. نتایج حاصل از این پژوهش می‌تواند در تدوین راهبردهای عملیاتی مبنی بر داده کاوی جهت پیش‌بینی فرار مالیاتی اصناف در استان‌ها مورد بهره برداری قرار گیرد.

واژه‌های کلیدی: فرار مالیاتی، مالیات اصناف، پیش‌بینی، داده کاوی

تاریخ دریافت: ۱۴۰۰/۰۹/۲۱ ، تاریخ پذیرش: ۱۴۰۱/۱۲/۲۲

۱. دانشجوی دکتری گروه مدیریت صنعتی، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران، mhm.ghasemi@yahoo.com
۲. استادیارگروه مدیریت صنعتی، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران، (نویسنده مسئول)، Abedi.sadegh@gmail.com
۳. دانشیار گروه مدیریت صنعتی، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران، alimohtashami07@gmail.com

doi 10.52547/taxjournal.30.56.1
[Downloaded from taxjournal.ir on 2026-11-07]
[DOI: 10.52547/taxjournal.30.56.1]

مقدمه

مالیات مهم ترین منبع مالی برای تأمین درآمدهای عمومی و یکی از مؤثرترین ابزارهای سیاست مالی در دنیاست که دولت به واسطه آن بسیاری از خدمات اجتماعی و رفاهی را در خدمت مردم قرار می‌دهد [24] (Yamen et al., 2018: 32). اتکا به درآمدهای مالیاتی منبع درآمدی پایدار و درون زای ملی را فرا روی دولت قرار می‌دهد [3] (Dehghani et al., 2019: 53). همواره تلاش برای بهره‌گیری از ابزارهایی که سود یا درآمد شرکت را کم و هزینه‌های شرکت را بیش از واقع نمایش می‌دهد وجود دارد تا مالیات کمتری پرداخت شود [۲۵] (Abdixhiku et al., 2017: 41). اجتناب مالیاتی از جمله مشکلاتی است که از قوانین و مقررات مالیاتی نشأت می‌گیرد و این امر در میان مؤدیان مالیات بر درآمد شرکت‌ها متداول‌تر است، زیرا بخش بزرگی از درآمدهای مالیاتی دولت را تشکیل می‌دهد [27] (Khalil et al., 2020: 40). در حال حاضر ممیزان مالیاتی با چالش شناسایی و جمع آوری مالیات از افرادی روبرو می‌باشند که به طور موفقیت آمیزی از پرداخت مالیات صحیح فرار می‌کنند. به منظور رویارویی با این چالش ممیزان مالیاتی به منابع محدود و راهبردهای حسابرسی سنتی تجهیز شده‌اند که زمانبر می‌باشد [28] (Agarwal et al., 2020: 188). در این خصوص دولت و مقامات مالیاتی، مجبور به تحمل هزینه‌های کشف و جلوگیری از فرار مالیاتی می‌شوند. به این دلیل همیشه پیدا کردن راه‌ها و شیوه هایی که بتوانند فرار مالیاتی را شناسایی و کشف کنند، مسئله‌ای مهم و چالش برانگیز بوده است. اگر دولت نتواند فرار مالیاتی را کشف کند، سرمایه گذاری‌های عمومی به دلیل کاهش مالیات‌ها و در پی آن کاهش درآمدهای دولت به طور منفی تحت تأثیر قرار خواهد گرفت [29] (Di Gioacchino et al., 2020: 62). آمار و تصاویر ایجاد شده در یک نظام مالیاتی کار آمد و عادلانه می‌تواند خود یک کنترل کننده نرم افزاری و علمی جهت ارتقای سطح کار آمدی اقتصادی و شفافیت اطلاعات منتشر شده در حوزه‌های مختلف اقتصادی هر کشور باشد [10] (Karimi et al., 2016: 23)، افزون بر این، در بسیاری از کشورهای در حال توسعه از جمله ایران، به دلیل نبود اطلاعات کافی و ضعف کارایی نظام اجرایی وصول مالیات، بسیاری از مشاغل پردرآمد از پرداخت مالیات متناسب با درآمدشان می‌گریزند و بخش قابل توجهی از درآمد دولت به این ترتیب از بین می‌رود [21] (Moghimi, 2009: 16). این رفتار مؤدیان مالیاتی در پرداخت نکردن مالیات را می‌توان به دو طبقه اجتناب مالیاتی و فرار مالیاتی تقسیم کرد. ۱۵ گروه صنفی در کشور عبارتند از وکلا، مشاوران حقوقی و خانواده، دفتر اسناد رسمی، مشاورین املاک، تالارهای پذیرایی، رستوران‌ها و کبابی‌ها، اغذیه‌فروشی‌ها و فست‌فودها، مراکز اقامتی اعم از هتل، هتل‌آپارتمان‌ها، مسافرخانه‌ها، نمایشگاه‌ها و فروشگاه‌های وسایل نقلیه، آرایشگاه‌های زنانه و مردانه، آجیل و خشکبار، قنادی، میوه و تره‌بار، مجموعه‌های فرهنگی ورزشی، لوازم‌التحریر، بازی‌های رایانه‌ای، دفاتر خدمات ارتباطی و دیگر اماکن در کل کشور که باید تحت نظارت سازمان امور مالیاتی قرار بگیرد [22] (Arab Mazar

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

& Dehghani, 2009: 17). فقدان پیش‌بینی علمی و برآورد توان مالیاتی افراد موجب شده است تا پیش‌بینی بر اساس روال سنواتی صورت گیرد . استفاده از ابزارها و روش‌های پیش‌بینی سنتی خطای بالایی دارد و به عملکرد ضعیف تری می‌انجامد [23] (Khosro Piraei et al., 2009: 16). با پیشرفت فناوری و علم تحلی داده، ابزارهای مدرنی به منظور یافتن موارد عدم رعایت و عدم تطبیق درآمدهای گزارش شده با سایر منابع به وجود آمده است [30] (Demir et al., 2020: 144). داده کاوی ابزارهای مفیدی جهت ارزیابی مالیاتی افراد ارائه می‌دهد. مساله اصلی در این پژوهش این می‌باشد که چگونه می‌توان از فرار مالیاتی اصناف جلوگیری نمود و آیا می‌توان با استفاده از تحلیل داده‌های سری زمانی مؤدیان مالیاتی اصناف، الگوی پیش‌بینی فرار مالیاتی اصناف را ارایه نمود. لذا با توجه به مساله این پژوهش، هدف این است که پس از پیش پردازش داده‌ها با استفاده از تکنیک‌های داده کاوی، مدل داده‌ای طراحی شود تا در ابتدا ویژگی‌های ورودی و خروجی جهت کلاس‌بندی داده‌های مؤثر بر رفتارهای مؤدیان مالیاتی در حوزه اصناف شناسایی شود و سپس با بهره‌گیری از آن، مؤدیان را طبقه‌بندی کرد تا رفتار آنها را بررسی نمود. بر اساس جستجوی بعمل آمده، بیشتر پژوهش‌های پیشن در حوزه فرار مالیاتی از رویکردهای تحلیل آماری و آزمون فرضیه استفاده شده است و این پژوهش برای اولین بار در حوزه امور مالیاتی استان قزوین در حوزه بررسی رفتار عملکرد مالیاتی اصناف اجرا می‌شود. ساختار اجرایی این پژوهش بدین صورت می‌باشد که در ابتدا با بررسی پژوهش‌های داخلی و خارجی در حوزه فرار مالیاتی، خلاءهای پژوهشی در خصوص امور مالیاتی اصناف شناسایی می‌گردد. سپس بر اساس روش پژوهش تدوین شده و انتخاب پایگاه اطلاعات سازمان امور مالیاتی استان قزوین در یک بازه زمانی ۶ ساله، پرونده‌های امور مالیاتی اصناف منتخب مورد بررسی قرار می‌گیرد. چهار مرحله در حوزه داده کاوی شامل یکپارچه‌سازی داده، تجمیع داده ها، نمونه‌گیری و پاکسازی داده‌ها به عنوان پیش پردازش داده‌ها انجام و سپس فرایند پاکسازی داده‌ها انجام خواهد گرفت. پس از انتخاب ویژگی‌های منتخب مؤدیان مالیاتی در حوزه اصناف با استفاده از الگوریتم داده کاوی شامل درخت تصمیم مدل پیش‌بینی فرار مالیاتی شناسایی و با استفاده از پارامترهای استاندارد، آنالیز حساسیت عملکرد مدل انجام خواهد گرفت . نهایتا پس از اعتبارسنجی مدل، نتایج مورد تجزیه و تحلیل قرار می‌گیرد. در انتها بر اساس نتایج به دست آمده پیشنهادات کاربردی برای سازمان امور مالیاتی و نیز به سایر پژوهشگران ارائه می‌گردد.

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

مبانی نظری

فرار مالیاتی

هرگونه تلاش غیر قانونی برای پرداخت نکردن مالیات مانند ندادن اطلاعات لازم در مورد عواید و منافع مشمول مالیات به مقامات مسئول، فرار مالیاتی نامیده می‌شود [31] (Yousefi et al., 2020: 75). در فرار مالیاتی، به رغم تطبیق فعالیت با قوانین، در اجرای فعالیت یک یا چند ضابطه مصوب مراجع ذی ربط تعمدا نادیده گرفته می‌شود [20] (Panahi , 2017: 1). وقتی مؤدی از گزارش درآمدهای مشمول مالیات خویش عدول کند یا به شیوه‌های عامدانه‌ای نظیر گزارش کمتر از واقع درآمد و فروش، گزارش بیش از واقع کسورات و یا تنظیم اظهارنامه غیرواقعی از تعهدات مالیاتی خود طفره رود، فرار مالیاتی محسوب می‌شود . روان و همکاران فرار مالیاتی را به عنوان کاهش غیرقانونی مالیات پرداختی با عدم گزارش درآمد با اعلام کردن نرخ‌های تفریق بالاتر تعریف می‌کند [32] (Ruan et al., 2019: 447). تمایز مفهومی بین فرار مالیاتی و اجتناب از پرداخت مالیات به قانونی یا غیر قانونی بودن رفتار مؤدیان مربوط می‌شود [33] (Hung , 2017: 52). وقتی که یک مؤدی مالیاتی از ارائه گزارش درست در مورد درآمدهای حاصل از کار با سرمایه خود که مشمول پرداخت مالیات می‌شود، امتناع می‌کند، یک نوع عمل غیر رسمی انجام می‌دهد که او را از چشم مقامات دولتی و مالیاتی کشور دور نگه می‌دارد [34] (Didi-mo et al., 2018: 110) . اما در اجتناب از مالیات، فرد نگران نیست که عمل او افشا شود . اجتناب از مالیات، از خلاهای قانونی در قانون مالیات‌ها نشات می‌گیرد. در اینجا فرد به منظور کاهش قابلیت پرداخت مالیات، خود دنبال راه‌های گریز می‌گردد [36] (Assylbekov et al., 2016: 56). مثلا درآمدهای نیروی کار را در قالب در آمد سرمایه نشان می‌دهد که از نرخ پایین تری برای مالیات برخوردارند. در اجتناب از مالیات، مؤدی مالیاتی دلیلی ندارد نگران احتمال افشا شدن باشد، چرا که او الزاما تمامی مبادلات خود را با جزئیات آن البته به شکل غیر واقعی، یادداشت و ثبت می‌کند [37] (Chen et al., 2010: 37). زمینه‌های پیدایش فرار مالیاتی بررسی‌های مختلف نشان می‌دهد شایع ترین زمینه‌های پیدایش فرار مالیاتی چنین است [38] (Devereux et al., 2002: 35):

۱. عدم گسترش فرهنگ مالیاتی در جامعه: فرهنگ مالیاتی، مجموعه‌ای از طرز تلقی، بینش و عکس العمل افراد در قبال نظام مالیاتی است. به عبارت دیگر طرز تلقی، بینش، برداشت، آرمان ها، ارزش‌های اجتماعی، قوانین جاری و میزان تحصیلات و آگاهی، از جمله عواملی هستند که فرهنگ مالیاتی را شکل می‌دهند [39] (Ferrantino et al., 2012:86)

۲. عدم مبادله کامل اطلاعات و نبودن سیستم نظارت و پیگیری در اخذ مالیات [40] (Goumagias et al., 2012: 53)

۳. تشخیص علی‌الراس و وجود ضعف در اجرای آن

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

۴. عدم استقبال از تسلیم اظهارنامه‌های مالیاتی و ضعیف بودن ضمانت‌های اجرایی: ارسال نکردن اظهارنامه، دلایل گوناگونی دارد. رایج ترین دلیل عدم ارسال، می‌تواند این باشد که در صورت به روز نبودن ثبت مؤدی، بهترین امکان برای ارسال نکردن اظهارنامه و پس از آن فرار از پرداخت مالیات فراهم شده است. دلیل دیگر عدم برخورد قاطع ماموران مالیاتی است که آن هم می‌تواند به دلیل عدم آشنایی و عدم تسلط کافی ماموران مالیاتی به قوانین مالیاتی و نداشتن مهارت کافی در اجرای آنها باشد

۵. تاخیر در وصول مالیات [42] (Jia et al., 2016: 13)

۶. عدم شناخت مؤدیان و مستند نبودن میزان درآمد آنها [44](Tian et al., 2016: 28)

۷. وجود معافیت‌های وسیع و متنوع: دولت‌ها گاه با اتخاذ برخی سیاست‌های حمایتی مالیاتی و معافیت‌های مالیاتی، راه فرار مالیاتی را باز می‌کنند و در صورتی که در سیستم اطلاعات کار آمد نیز در نظام مالیاتی موجود نباشد، فرار مالیاتی گسترش می‌یابد. همچنین با اعطای معافیت‌های مالیاتی، گروهی از پرداخت مالیات معاف می‌شوند و این امر به کاهش درآمدهای مالیاتی دولت منجر می‌شود. بنابراین، اگر اعطای معافیت‌ها با دقت لازم صورت نگیرد، دولت برای کسب درآمد مالیاتی مورد نظر، فشار سنگینی را بر گروه‌های دیگر جامعه وارد می‌کند. یعنی برای جبران کمبود درآمد مالیاتی، میزان نرخ‌های مالیاتی بیشتری، بر گروه هایی که از معافیت‌های مالیاتی بهره مند نیستند، تحمیل خواهد شد. بی تردید اعطای معافیت‌های مالیاتی گسترده، به ناکارایی نظام مالیاتی و ضعف نظام اجرایی مالیاتی منجر می‌شود [41] (Hsu et al., 2015: 89).

پیشینه پژوهش

در این راستا می‌توان به پژوهش‌های داخل کشور اشاره نمود. رضاقلی زاده و همکاران (۱۳۹۸) به بررسی تحلیل فرار مالیاتی در ایران به روش شاخص چندگانه [1] (Rezaqoli et al., 2019: 26)، خداپرست و همکاران (۱۳۹۷) به بررسی برآورد سهم نسبی عوامل مدیریتی بر فرار مالیاتی [2] (Khodaparast et al., 2019: 10)، محمدی (۱۳۹۷) به بررسی تأثیر فرار مالیاتی و فساد بر کارآفرینی مورد مطالعه قرار دادند [4] (Mohammadi, 2019: 11). حمیدی و همکاران (۱۳۹۴) به بررسی تشخیص فرار مالیاتی مؤدیان مالیات بر ارزش افزوده با استفاده از الگوریتم‌های طبقه‌بندی و خوشه‌بندی [11] (Hamidi et al., 2016: 23)، فطرس و همکاران(۱۳۹۵) به بررسی اقتصاد زیرزمینی و فرار مالیاتی در چارچوب مدل‌های تعادل عمومی پویای تصادفی پرداختند [8] (Fetras et al., 2017: 7). در زمینه مورد مطالعه در استان‌های داخلی می‌توان به پژوهش‌های کرمخانی و همکاران(۱۳۹۵) بررسی میزان کارایی جرایم مالیاتی در جلوگیری از فرار مالیاتی در نظام مالیات بر ارزش افزوده در بین مؤدیان مالیاتی استان ایلام [9] (Karamkhani et al., 2017: 2)، لاری دشت بیاض، قائم مقامی و

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

کهرمی (۱۳۹۵) به بررسی عوامل مؤثر بر فرار مالیاتی در استان خراسان جنوبی با تأکید بر مؤلفه‌های فرهنگی [15] (Lari Dasht et al., 2017: 1)، زهی نقی و همکاران(۱۳۸۹) به بررسی عوامل موثر بر فرار مالیاتی (مطالعه موردی استان آذربایجان شرقی) پرداختند [13] (Zahi et al., 2011: 18). نتایج این پژوهش نشان داد مالیات از اهم درآمدهای دولت است که پذیرش آن از طرف ملت و نحوه اجرای آن از طرف دولت، به ویژگی‌های فرهنگی هر کشور بستگی دارد. در زمینه بکارگیری روش‌های داده کاوی در تشخیص فرار مالیاتی می‌توان به مطالعات تقوی فرد و همکاران(۱۳۹۶) با موضوع تحلیل آینده نگر تشخیص فرار مالیاتی مؤدیان مالیات بر ارزش افزوده [14] (TaqviFard et al., 2018: 25) و نیز مطالعات مداح و همکاران (۱۳۹۴) با موضوع بررسی فرار مالیاتی در واردات ایران [12] (Maddah et al., 2016: 20)، رحیمی کیا و همکاران (۱۳۹۴) با عنوان تشخیص فرار مالیاتی با استفاده از سیستم هوشمند ترکیبی [16] (Rahimi Kia et al., 2016: 23)، دستگیر و همکاران (۱۳۹۴) با موضوع کاربست روش‌های داده کاوی به منظور ارتقای عملکرد تشخیص فرار مالیاتی اشاره نمود [17] (Tasgir et al., 2016: 23).

در زمینه پژوهش‌های خارج کشور که در چند سال اخیر به چاپ رسیده است می‌توان به خلیل و همکاران (۲۰۲۰) تأثیر دینداری بر نگرش فرار مالیاتی در لبنان اشاره نمود [27] (Khalil et al., 2020: 40). دی گیوچینو و همکاران (۲۰۲۰) در بررسی فرار مالیاتی و روحیه مالیاتی پرداختند [28] (Di Gioacchino et al., 2020: 62). اثرات اخلاقی مالیات و هنجارهای اجتماعی بر فرار مالیاتی را در هنگام تعامل افراد در یک شبکه بررسی شده است. الگویی ارائه شده است که شامل انگیزه هایی برای رعایت مالیات به صورت مجازات و جریمه، روحیه مالیاتی و شهرت برای رفتار اجتماعی است. همچنین دمیر و همکاران (۲۰۲۰) در پژوهش‌های خود به بررسی تغییرات سیاست تجاری، فرار مالیاتی و قانون بنفورد [30] (Demir et al., 2020: 144)، روان و همکاران (۲۰۱۹) بررسی شناسایی گروه‌های مشکوک فرار مالیاتی مبتنی بر معاملات وابسته در داده‌های بزرگ [32] (Ruan et al., 2019: 447)، دیدیمو و همکاران (۲۰۱۸) بررسی یک سیستم تجزیه و تحلیل بصری برای پشتیبانی از کشف فرار مالیاتی پرداختند [34] (Di Didimo et al., 2018: 110). این پژوهش، یک سیستم پشتیبانی تصمیم برای کشف فرار مالیاتی، مبتنی بر یک زبان بصری قدرتمند و تکنیک‌های پیشرفته تجسم شبکه را توصیف می‌کند. با همکاری آژانس درآمد ایتالیا که در حال حاضر از آن استفاده می‌شود، توسعه یافته است. یامن (۲۰۱۸) در پژوهش خود تحت عنوان تاثیر کیفیت محیط سازمانی در فرار از پرداخت مالیات‌ها با هدف مقایسه مقایسه تأثیر کیفیت محیط سازمانی (IEQ) بر فرار مالیاتی در کشورهای عضو قدیمی (قبل از ۲۰۰۴) و جدید (پس از ۲۰۰۴) اتحادیه اروپا توسط شاخص‌های مدیریت جهانی بانک جهانی اندازه‌گیری صورت پذیرفت [24] (Yamen et al., 2018: 32). هانگ (۲۰۱۷) در مقاله خود به بررسی توضیح غیرخطی بودن تورم و رشد

[Downloaded from taxjournal.ir on 2026-11-07]
[DOI: 10.52547/taxjournal.30.56.1]

اقتصادی: نقش فرار مالیاتی پرداختند. این مقاله شامل فرار مالیاتی در یک مدل رشد درون زای ساده با پول در عملکرد اقتصادی است [33] (Hung , 2017: 52).

در مطالعات انجام شده خلا پژوهشی بررسی امور مالیاتی اصنافی می‌باشد که در سازمان بورس پذیرفته نشده‌اند و شفافیت لازم جهت مالیات ابرازی خود ندارد. تمامی مشاغل و اصناف باید مالیات خود را سالانه از طریق اظهارنامه مالیاتی پرداخت کنند. اما چگونگی محاسبه مالیاتی اصناف یکی از مهم ترین سوالاتی است که برای سازمان امور مالیاتی به وجود می‌آید. تاکنون پژوهش‌های اندکی در زمینه فرار مالیاتی در حوزه بررسی اصناف داخل کشور انجام شده است و اغلب پژوهش‌های انجام شده در این زمینه با توجه به رویکردهای اقتصادی به پیش‌بینی فرار مالیاتی در شرکت‌های بورسی پرداخته‌اند. با وجود این، حتی در پژوهش هایی که از منظر ارقام اقتصادی به مسئله فرار مالیاتی پرداخته‌اند، جای خالی استفاده از رویکردهای نوینی چون داده کاوی مشهود است. این مسئله بدان علت مهم است که در پژوهش‌های بسیاری سودمندی، کارایی و عملکرد بهتر تکنیک‌های داده کاوی در مقابل روش‌های خطی به وضوح مورد تائید قرار گرفته است.

روش پژوهش

این پژوهش از لحاظ هدف از نوع توصیفی- پیمایشی بوده و در آن جهت جمع آوری اطلاعات و داده‌های لازم، از پایگاه داده امور مالیاتی استان قزوین به عنوان یک مولفه ی بازیابی اطلاعات استفاده شده است. همچنین داده‌ها از چند منبع مختلف در ادارات مختلف استانی ترکیب شده و با استفاده از روند استخراج، پالایش، بارگذاری (ETL) و یکپارچه شده‌اند. بدین منظور این پژوهش در دو فاز اساسی صورت گرفت. در فاز اول این پژوهش ابتدا ویژگی‌های مختلف را برای فرار مالیاتی اصناف تعریف نموده تا از طریق آن و با استفاده از روش‌های داده کاری به یک تقسیم‌بندی کلی در زمینه نمونه‌های آماری دست آوریم. در فاز دوم پژوهش برای تقسیم‌بندی اصناف به دسته‌های مختلف از بانک اطلاعات سازمان امور مالیاتی استان قزوین استفاده گردید و سپس با توجه به آن تحلیل‌های مختلف صورت گرفت.

در این پژوهش جهت هم مقیاس‌سازی داده به دلیل آنکه از پایگاه‌های داده مختلف انتخاب شده بودند ابتدا متغیرهای تحلیلی را انتخاب سپس آنها را هم مقیاس نمؤدیم. چهار مرحله در پیش پردازش داده در داده کاوی عبارت است از

- یکپارچه‌سازی داده (Data Integration)
- تجمیع داده (Data Aggregation)
- نمونه‌گیری (Data Sampling)

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

-پاک‌سازی داده‌ها (Data Cleaning)

- انتخاب زیر مجموعه‌ای از ویژگی‌ها (Feature Subset selection)

-تغییر شکل متغیرها (Variable Transformation)

لازم به ذکر است مرحله ۵ و ۶ را می‌توان جز فرایندهای مرحله پاک‌سازی داده‌ها نیز در نظر گرفت.در این پژوهش از روش اتصال محکم[1] جهت یکپارچه‌سازی داده‌ها استفاده شده است. بدین صورت که از پایگاه داده امور مالیاتی استان قزوین به عنوان یک مولفه بازیابی اطلاعات استفاده شده است. همچنین داده‌ها از چند منبع مختلف در ادارات مختلف استانی ترکیب شده و با استفاده از روند استخراج، پالایش، بارگذاری (ETL) و یکپارچه شده‌اند. جهت هم مقیاس‌سازی داده به دلیل آنکه از پایگاه‌های داده مختلف انتخاب شده بودند ابتدا متغیرهای تحلیلی را انتخاب سپس آنها را هم مقیاس نمؤدیم. تعریف ویژگی‌های اولیه پژوهش بر اساس پایگاه داده در سازمان عبارت است از:

- واحد مالیاتی (بر اساس تعداد پرونده مالیاتی)
- سابقه فعالیت (سال)
- جنسیت (زن و مرد)
- کد ملی (بر اساس کد ۱۰ رقمی)
- تسلیم اظهارنامه (بر اساس صفر و یک می‌باشد)
- نوع اتحادیه‌ها (قانون سابق)
- مالیات ابرازی(میزان مالیات اعلامی توسط مؤدی بر اساس ریال)
- مالیات قطعی (محاسبه مالیات قطعی شده بر اساس ریال)
- مبلغ فرار مالیاتی (بر حسب ریال)
- طبقه‌بندی فرار مالیاتی (براساس دو کلاس بلی و خیر)

بر اساس تقسیم‌بندی قدیم اتحادیه‌ها در سنوات گذشته، به ۱۵ بخش تقسیم‌بندی می‌شوند. در این تحقیق ۵۶۰۰ پرونده مالیاتی مرتبط با اصناف در پنج گروه مالیاتی به عنوان داده‌های نمونه انتخاب گردید. به دلیل پیچیدگی محاسبات تحلیل داده‌های ۱۵ صنف طی ۶ سال(بالای ۱۶۰۰۰ پرونده)، وجود داده‌های ناقص و عدم خروجی قابل قبول در پیش داده کاوی، لذا از کل جامعه آماری، نمونه‌گیری در قالب ۵ گروه صنفی صورت پذیرفت. این پنج گروه عبارت است از :

- G1 :گروه صنفی صاحبان دفاتر رسمی و حقوقی و مشاغل وابسته

1 .Tight Coupling

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

- G2 : گروه صنفی مشاورین املاک
- G3 : گروه صنفی تالارهای پذیرایی، رستوان و مشاغل وابسته
- G4 : گروه صنفی خدمات ارتباطی و مشاغل وابسته
- G5 : گروه صنفی نمایشگاه و فروشگاه لوازم خودرویی و مشاغل وابسته

صاحبان مشاغل از ابتدای سال ۱۴۰۰ (برابر اصلاحیه آیین نامه اجرایی موضوع ماده ۹۵ قانون مالیات‌های مستقیم) بر مبنای نوع فعالیت و درآمد حاصل از فروش کالا و خدمات بدون در نظر گرفتن صنف به سه گروه مؤدیان گروه اول، دوم و سوم تقسیم‌بندی می‌شوند. طبق این قانون گروه سوم شامل صاحبان مشاغل و حرفی که درآمد ابرازی (فروش کالا یا فروش توأم کالا و ارائه خدمت) دو سال قبل آن‌ها و با آخرین درامد قطعی شده آن‌ها (فروش کالا یا فروش توأم کالا و ارائه خدمت) کمتر از هجده میلیارد ریال باشد (مشروط به ابلاغ برگ قطعی حداکثر تا پایان دی‌ماه سال قبل به شخص مؤدی یا نماینده قانونی وی) باشد. مطابق این تقسیم‌بندی تمامی ۵ گروه صنفی منتخب در گروه سوم قرار می‌گیرند. مؤدیان گروه سوم: بر اساس آیین نامه اجرایی موضوع ماده ۹۵ اصلاحی قانون مالیات‌های مستقیم، مؤدیان گروه سوم، مؤدیانی هستند که دارای کسب و کارهای کوچک خرد(درآمد قطعی کمتر از ۱۰ میلیارد ریال) بوده و در شرایط دسته‌بندی مؤدیان گروه اول و دوم، قرار نمی گیرند. مطابق این تقسیم‌بندی تمامی ۵ گروه صنفی منتخب در گروه سوم قرار می‌گیرند. این ۵ گروه منتخب دارای ویژگی‌های مشابه‌ای می‌باشند. اولا این که همگی در گروه سوم تقسیم‌بندی جدید ۱۴۰۰ قرار می‌گیرند. ثانیا تمامی داده‌های ۶ ساله این پنج گروه صنفی در پایگاه داده استانی موجود و در اختیار قرار گرفت. آماری که در این مقاله ارائه شده است شامل آمار توافقات که مورد پذیرش سازمان قرار گرفته است نمی باشد. آمار ارایه شده شامل دو بخش می‌باشند مؤدیانی که اظهارنامه داده‌اند و مؤدیانی که اظهارنامه نداده‌اند. مؤدیانی که اظهارنامه داده‌اند نیز به دو دسته تقسیم‌بندی می‌شود. اظهارنامه آنها در قالب توافق راستی‌آزمایی شده است و مغایرت ندارد و یا مغایرت دارد . مؤدیانی که مغایرت ندارند و نیز مؤدیانی که توافقات آنها مورد پذیرش سازمان قرار گرفته است در آمار ارائه شده در این مقاله نمی باشد.بنابراین طبق بررسی‌های بعمل آمده حدود ۵۰ درصد از اظهارنامه‌ها مورد رسیدگی قرار می‌گیرند.

فرایند کاری بدین صورت می‌باشد پس از فراخوان ارایه اظهارنامه مالیاتی، مؤدیان وارد توافق یاتبصره ماده ۱۰۰ می‌شوند . سپس سازمان با استفاده از پایگاه‌های مالیاتی راست آزمایی کرده و سپس برگ تشخیص برای مؤدیانی که مغایرت در ابراز درآمد با پایگاههای اطلاعاتی دارند صادر می‌گردد. در ادامه پس از طی فرآیند دادرسی، برگ قطعی صادر می‌گردد که این مبنای فرار مالیاتی می‌باشد. داده ارائه شده شامل دو بخش می‌باشد :

بخش اول شامل مؤدیانی می‌باشد که اظهارنامه مالیاتی ارائه نداده‌اند و به تبع آن به تکالیف قانونی عمل

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

نکرده‌اند لذا سازمان امور مالیاتی بر اساس داده‌های موجود در پایگاه‌های در اختیار، مالیات آنها را تشخیص می‌دهد. پس در فرآیند رسیدگی حسابرسی مبتنی بر ریسک، حتی اگر آنها در سال بعد هم اظهارنامه ارائه نمایند قطعاً مؤدی پر ریسک طبقه‌بندی می‌شوند.

بخش دوم شامل مؤدیانی می‌باشند که اظهارنامه داده‌اند یا در فرآیند تبصره ۱۰۰ یا توافق مالیاتی قرار گرفته‌اند. با توجه به این که سازمان امور مالیاتی به جهت رونق کسب و کار، هر سال برای برخی از گروه‌های شغلی و یا صاحبان مشاغل تسهیلاتی در قالب تبصره ماده ۱۰۰ قانون مالیات‌های مستقیم در نظر می‌گیرد؛ به این صورت که هرگاه درآمد حاصل از فروش (کالا یا خدمات) برای این گروه در یک سال مالیاتی کمتر از سقف معین از سوی این سازمان باشد، از انجام بخشی از تکالیف مالیاتی معاف خواهند بود. لذا داده‌های مرتبط با مؤدیان مشمول تبصره ماده ۱۰۰ در تحلیل‌های این پژوهش در نظر گرفته نشده است. این مؤدیان در صورت انجام مراحل قانونی و اظهارنامه مالیاتی جز مؤدیان خوش حساب می‌باشند و جز کلاس‌بندی فرار مالیاتی قرار نخواهند گرفت.

لیکن سازمان با توجه به پایگاه‌های اطلاعاتی در اختیار دارد آنها را مورد راستی آزمایی قرار می‌دهد چنانچه مغایرتی بین درآمد ابرازی و درآمد محاسبه شده وجود داشته باشد باعث خروج از شرایط توافق باعث می‌شود و این مؤدی در فرآیند حسابرسی مبتنی ریسک قرار بگیرد . بنابراین ذکر این نکته ضروری است که اگر بین ابراز مؤدی و اطلاعات مستندی که سازمان در سامانه وجود دارد مغایرت وجود داشته باشد مصداق کتمان و فرار مالیاتی میباشد که قانون گذار در قانون ۱۹۲ جرایم و آن را اعلام نموده است .

پاک‌سازی داده‌ها

برای انجام یک داده کاوی مطلوب لازم است مقادیر گمشده جایگزین شوند. داده‌های مزاحم شناسایی و به نحوی مناسب با آنها برخورد و ناپایداری‌ها اصلاح شوند.جهت آماده‌سازی داده‌ها جهت مدل‌سازی نیاز است پاک‌سازی داده‌ها صورت پذیرد.

انتخاب پایگاه داده مناسب

بر اساس مراحل قبل ۵۴۰۰ پرونده مالیاتی مرتبط با اصناف در قالب یک پایگاه داده یکپارچه و تجمیع شده انتخاب گردید.

انتخاب زیر مجموعه‌ای از ویژگی‌ها

ویژگی‌های غیر مرتبط، ویژگی هایی که شامل هیچ گونه اطلاعات مفیدی برای انجام عملیات داده کاوی نیستند. ویژگی‌های زائد، حاوی اطلاعاتی هستند که همه یا بخش زیادی از آن در یک یا چند ویژگی دیگر تکرار شده باشد. در این گام از بین ویژگی‌های شناسایی شده در مرحله تجمیع داده‌ها ویژگی انتخاب گردید. لازم به ذکر است کد

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

مالیاتی، کد ملی، مالیات قطعی ویژگی هایی بودن که در فرایند مدل‌سازی تاثیر گذار نمی باشند، که حذف گردیدند.

جایگزین کردن داده‌های گم شده

مطابق تحلیل انجام شده، درآمد ابرازی شامل ۵۶ داده گم شده، مالیات ابرازی ۱ داده گم شده و فرار مالیاتی ۴۲ داده گم شده می‌باشند. لذا در این گام داده‌های گم شده را با میانگین داده‌ها جایگزین گردید.

فیلتر کردن داده‌ها

در این گام داده‌های مورد تحلیل را بر اساس گروه‌های صنفی منتخب پایش می‌نماییم . به عنوان نمونه در تحلیل داده‌های گروه صنفی صاحبان دفاتر رسمی و حقوقی و مشاغل وابسته مشاهده می‌شود. گروه صنفی G1 که سابقه فعالیت بالای ۱۴ سال را دارند فرار مالیاتی ثبت نشده است.

مشخص کردن برچسب/لیبل

در این گام می‌بایست یکی از ویژگی‌های موجود در پایگاه داده را به عنوان متغیر پاسخ انتخاب نماییم. لذا ویژگی فرار مالیاتی را بر اساس دو کلاس فرار مالیاتی دارد (Yes)، فرار مالیاتی ندارد(No) انتخاب می‌نماییم.

جهت آستانه پذیرش فرار مالیاتی از ماده ۱۹۴ قانون مالیاتی مبنای اختلاف بین مالیات ابرازی و برگ قطعی می‌باشد. طبق این ماده قانونی مؤدیانی که اظهارنامه آن‌ها در اجرای مقررات ماده (۱۵۸) این قانون مورد رسیدگی قرار می‌گیرد در صورت که درآمد مشمول مالیات مشخصه قطعی با رقم اظهار شده از طرف مؤدی بیش از پانزده درصد (۱۵٪) اختلاف داشته باشد علاوه بر تعلق جرائم مقرر مربوط که قابل بخشودن نیز نخواهد بود تا سه سال بعد از ابلاغ مالیات مشخصه قطعی از هر گونه تسهیلات و بخشودگی‌های مقرر در قانون مالیاتها نیز محروم خواهند شد.

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

مدل‌سازی و تحلیل داده‌ها

پس از مرحله آماده‌سازی و پیش پردازش داده‌های فرار مالیاتی اصناف، در این پژوهش از الگوریتم داده کاوی شامل درخت تصمیم استفاده می‌شود. یکی از پرکاربردترین الگوریتم‌های داده کاوی، الگوریتم درخت تصمیم است. در داده کاوی، درخت تصمیم یک مدل پیش‌بینی کننده است به طوری که می‌تواند برای هر دو مدل رگرسیون و طبقه‌ای مورد استفاده قرار گیرد. زمانی که درخت برای کارهای طبقه‌بندی استفاده می‌شود، به عنوان درخت طبقه‌بندی[1] شناخته می‌شود و هنگامی که برای فعالیت‌های رگرسیونی به کار می‌رود درخت رگرسیون[2] نامیده می‌شود. در ساختار درخت تصمیم، پیش‌بینی به دست آمده از درخت در قالب یک سری قواعد توضیح داده میشود. هر مسیر از ریشه تا یک برگ درخت تصمیم، یک قانون را بیان میکند و در نهایت برگ با کلاسی که بیشترین مقدار رکورد در آن تعلق گرفته برچسب می‌خورد. پارامترهای درخت تصمیم فرار مالیاتی اصناف در جدول زیر مشاهده می‌شود .

جدول (۱)ـ پارامترهای درخت تصمیم فرار مالیاتی اصناف

پارامتر	مقدار	توضیحات
حداکثر عمق	۲۰	عمق یک درخت بسته به اندازه و ماهیت مجموعه نمونه متفاوت است. این پارامتر برای محدود کردن اندازه درخت تصمیم استفاده می‌شود.
میزان اطمینان	۰/۲۵	این پارامتر میزان اطمینان مورد استفاده برای محاسبه خطای بدبینانه هرس را مشخص می‌کند.
حداقل سود	۰/۱	سود یک گره قبل از تقسیم آن محاسبه می‌شود.
حداقل اندازه برگ	۲	درخت به گونه‌ای تولید می‌شود که هر زیر مجموعه برگ حداقل تعداد نمونه برگ حداقل را داشته باشد.
حداقل اندازه برای تقسیم	۴	فقط آن گره‌ها تقسیم می‌شوند که اندازه آنها بزرگتر یا مساوی حداقل اندازه پارامتر split باشد.
تعداد گزینه‌های پیش از هرس	۳	این پارامتر تعداد گره‌های جایگزین را که برای تقسیم انتخاب می‌شود را تنظیم می‌کند.

منبع: یافته‌های محقق

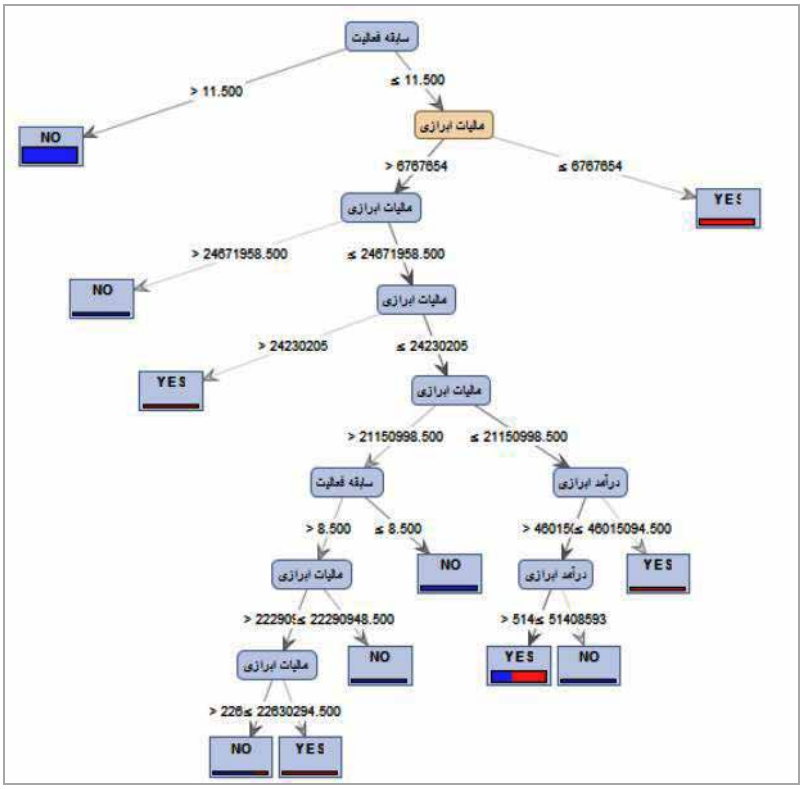
در شکل زیر درخت تصمیم استخراج شده مشاهده می‌شود:

1. Classification Tree
2 . Regression Decision Tree

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

شکل (۱)ـ درخت تصمیم فرار مالیاتی اصناف



بررسی عملکرد مدل پیش‌بینی فرار مالیاتی اصناف

در این قسمت با توجه به انتخاب الگوریتم مدل ساز درخت، مدل پیش‌بینی را طراحی می‌نمایم. مدل پیش‌بینی بر اساس داده‌های آموزشی، پایگاه داده جدید را برای پیش‌بینی بدست می‌آوریم. هدف این می‌باشد بر اساس داده‌های آموزشی و یادگیری ماشین، مدل بتواند رفتار داده‌های جدید را پیش‌بینی نماید. پس از تفکیک داده‌ها به داده‌های آموزشی و داده‌های تست می‌توانیم، صحت ودقت عملکرد مدل پیش‌بینی را اندازه‌گیری نماییم. بر اساس پیش مدل‌سازی انجام شده، نسبت به پاک‌سازی داده‌ها اقدام نموده و سپس داده‌ها را با نسبت معمول ۷۰ به ۳۰ به داده‌های آموزشی و داده‌های تست تقسیم‌بندی می‌نماییم. این بدین معنا می‌باشد که ۷۰ درصد داده ها، یادگیری الگوریتم را انجام می‌دهند و ۳۰ درصد از آنها الگوریتم ارایه شده را تست نمایند.

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

یکی از این آزمون‌ها جهت صحت مدل در حوزه هوش مصنوعی، ماتریس آشفتگی[1] می‌باشد که در آن عملکرد مدل را نشان می‌دهند. بر اساس این ماتریس مشاهده می‌شود، مدل ساز بر اساس دو کلاس فرار مالیاتی دارد(YES) و فرار مالیاتی ندارد(NO)، ۶۷۲ نمونه که در کلاس فرار مالیاتی ندارد(NO) را درست پیش‌بینی کرده است. لذا ۶۰ درصد پیش‌بینی در کلاس فرار مالیاتی ندارد(NO) را مدل ساز به درستی پیش‌بینی کرده است. که نشان از صحت و دقت بالا در پیش‌بینی دارد. هنچنین ۴۰۲ نمونه که در کلاس فرار مالیاتی دارد (YES) را درست پیش‌بینی کرده است. لذا ۸۳ درصد پیش‌بینی در کلاس فرار مالیاتی دارد (YES) را مدل ساز به درستی پیش‌بینی کرده است که نشان از صحت و دقت بالا در پیش‌بینی دارد.

جدول (۲)ـ ماتریس آشفتگی مدل‌ساز مدل فرار مالیاتی اصناف برا ی داده‌های آموزشی

Accuracy: 66.83%			
پیش‌بینی کلاس	پیش‌بینی درست (Yes)	پیش‌بینی درست (NO)	
٪۶۰٫۰۵	۴۴۷	۶۷۲	کلاس فرار مالیاتی ندارد (NO)
٪۸۲٫۳۸	۴۰۲	۸۶	کلاس فرار مالیاتی دارد (Yes)
	٪۴۷٫۳۵	٪۸۸٫۶۵	Recall

منبع: یافته‌های محقق

آنالیز حساسیت عملکرد مدل پیش‌بینی

ارزیابی عملکرد الگوریتم‌های شرح داده شده در بالا، با استفاده از معیارهای متفاوتی بر مبنای دیدگاه حساسیت و تشخیص صورت گرفته شده‌است. حساسیت و تشخیص دو شاخص برای ارزیابی نتیجه یک دسته‌بندی دودویی (دوحالته) هستند. زمانی که بتوان داده‌ها را به دو گروه مثبت و منفی تقسیم کرد، دقت نتایج یک آزمایش که اطلاعات را به این دو دسته تقسیم می‌کند با استفاده از شاخص‌های حساسیت و ویژگی قابل اندازه‌گیری و توصیف است.

- معیار صحت (Accuracy) نسبت مقداری موارد صحیح طبقه‌بندی‌شده توسط الگوریتم از یک کلاس مشخص، به کل تعداد مواردی که الگوریتم چه به‌صورت صحیح و چه به‌صورت غلط، در آن کلاس طبقه‌بندی کرده است را محاسبه می‌کند. که در این پژوهش معیار صحت برابر ۶۷ درصد می‌باشد.
- معیارِ کاربردیِ دیگری به نام پوشش (Recall) نیز وجود دارد و همان‌طور که از نامش پیداست، به

1. Confusion Matrix

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

دنبال محاسبه‌ی پوشش بر روی کل داده‌هاست. همان‌طور که مشاهده می‌کنید، تمرکز اصلی معیار پوشش بر خلاف معیار صحت بر روی داده‌هایی است که واقعاً «YES» بوده‌اند. که در این پژوهش معیار پوشش برابر ۶۸ درصد می‌باشد. در برخی از مسائل، معیار پوشش مهم‌تر از معیار صحت می‌شود. به معیار پوشش، معیار حساسیت[1] نیز گفته می‌شود.

- معیار کاپا (Kappa) به می‌تواند عملکرد الگوریتم‌های چند کلاسه را ارزیابی کند. این معیار قادر است که تعداد نمونه‌ها را در یک مجموعه‌ی داده‌ی نامتوازن نیز در نظر بگیرد و کیفیت الگوریتم را با توجه به تعداد نمونه‌ها بیان کند. پس برای مسائلی با طبقه‌های مختلف و همچنین مسائلی که داده‌های نامتوازن دارند، می‌توانید از این معیار استفاده نماییم. معیار کاپا مدل برابر ۰٫۶۱۲ می‌باشد که عملکرد خوب(بین ۰٫۶ تا ۰٫۸) مدل ساز را نشان می‌دهد.
- ضریب همبستگی کندال (Kendall tau) که به تای کندال مشهور است برای سنجش قدرت رابطه میان دو دو متغیر کلاس‌بندی واقعی و پیش‌بینی شده به کار می‌رود. این آزمون نیز برای سنجش میزان پیوستگی میان دو متغیر استفاده می‌شود. تای کندال همیشه بین ۱ و ۱- است که این دو مقدار متناظر با تطابق و عدم‌تطابق کامل است. انتظار می‌رود تای کندال برای دو متغیر مستقل نزدیک به صفر باشد. ضریب هماهنگی کندال نیز مانند ضریب همبستگی اسپیرمن، به جای مقدار از ترتیب مقدارها برای اندازه‌گیری میزان وابستگی استفاده می‌کند. ضریب همبستگی رتبه‌ای کندال در این پژوهش برابر ۰٫۶۳۲ می‌باشد که عملکرد خوب(بین ۰٫۶ تا ۰٫۸) مدل ساز را نشان می‌دهد.در جدول زیر مقادیر پارامترهای تحلیل حساسیت مدل مشاهده می‌شود.

1 . Sensitivity

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

جدول (۳)ـ پارامترهای تحلیل حساسیت مدل

پارامتر لاتین	پارامتر	مقدار	توضیحات
Accuracy	دقت	٪۶۶٫۸۳	درصد پیش‌بینی‌های صحیح
Kendall tau	ضریب کندال تاو	۰٫۶۳۲	قدرت رابطه بین دو متغیر کلاس‌بندی واقعی و پیش‌بینی شده
Kappa	ضریب کاپا	۰٫۶۱۲	ارزیابی کیفیت الگوریتم‌های طبقه‌بندی چند کلاسه
Weighted Mean Recall	ضریب پوشش میانگین وزنی	٪۶۸٫۰۰	محاسبه‌ی پوشش بر روی کل داده ها
Weighted Mean Precision	دقت وزنی متوسط	٪۷۱٫۲۲	میانگین دقت وزنی در هر کلاس مجزا
Spearman rho	ضریب اسپیرمن	۰٫۶۳۲	همبستگی رتبه‌ای بین پیش‌بینی کلاس‌بندی و واقعی
Correlation	ضریب‌همبستگی	۰٫۶۳۲	ضریب‌همبستگی
Cross_Entropy	آنتروپی متقاطع	۰٫۴۸۱	بررسی میزان آشفتگی و بی نظمی در کلاس بندی

منبع: یافته‌های محقق

در جدول زیر مقادیر پارامترهای خطای مدل مشاهده می‌شود

جدول (۴)ـ پارامترهای تحلیل خطای مدل

پارامتر لاتین	پارامتر	مقدار	توضیحات
Classification Error	خطای طبقه بندی	٪۳۳٫۱۷	درصد پیش‌بینی‌های نادرست
Absolute Error	خطای مطلق	۰٫۱۱۶ +/- ۰٫۰۹۷	میانگین قدرمطلق پیش‌بینی از مقدار واقعی
Relative Error	خطای نسبی	٪۱۱٫۵۶ +/- ٪۳٫۶۹	میانگین خطای نسبی
Squared Error	خطای مربعات	۰٫۲۱۱ +/- ۰٫۱۷۹	میانگین خطای مربعات

منبع: یافته‌های محقق

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

اعتبارسنجی مدل بر اساس داده‌های تست (آزمون)

برای اعتبارسنجی مدل، داده‌ها را به دو بخش داده‌های آموزش و تست به نسبت ۷۰ به ۳۰ تقسیم تقسیم‌بندی شد. اما نمی دانیم داده هایی که برای تست استفاده شده است چقدر شبیه داده می‌باشند که برای آموزش و ساخت مدل استفاده شده‌اند. برای این منظور از تکنیک Cross Validation، صحت اعتبار سنجی را مورد آزمون قرار دهیم تا با اطمینان بیشتری بتوانیم درصد عملکرد مدل‌سازی را تخمین بزنیم و مطمین شویم داده هایی که برای تست استفاده شده است مشابه داده‌های آموزشی می‌باشند. در جدول زیر صحت اعتبار سنجی مدل پیش‌بینی را با استفاده از تکنیک Cross Validation، مورد آزمون قرار داده شده است. تا با اطمینان بیشتری بتوانیم درصد عملکرد مدل‌سازی را تخمین بزنیم.

جدول (۵)- اعتبارسنجی بر اساس داده‌های آزمون

Accuracy: 67.79% +/- 1.99%		
پیش‌بینی کلاس	پیش‌بینی درست(Yes)	پیش‌بینی درست(NO)	
٪۶۱٫۶۷	۱۳۳۷	۲۱۵۱	کلاس فرار مالیاتی ندارد (NO)
٪۷۹٫۲۰	۱۴۸۱	۳۸۹	کلاس فرار مالیاتی دارد (Yes)
	٪۵۲٫۵۶	٪۸۴٫۶۹	Recall

منبع: یافته‌های محقق

معیار صحت (Accuracy) برابر ٪۶۷٫۷۹ +/- ٪۱٫۹۹ نشان از قابلیت اطمینان و اعتبار مناسب جهت مدل پیش‌بینی می‌باشد . به دلیل تکرارهای مختلف جهت انتخاب داده‌های آموزشی +/- ٪۱٫۹۹ انحراف معیار محاسبه شده است.

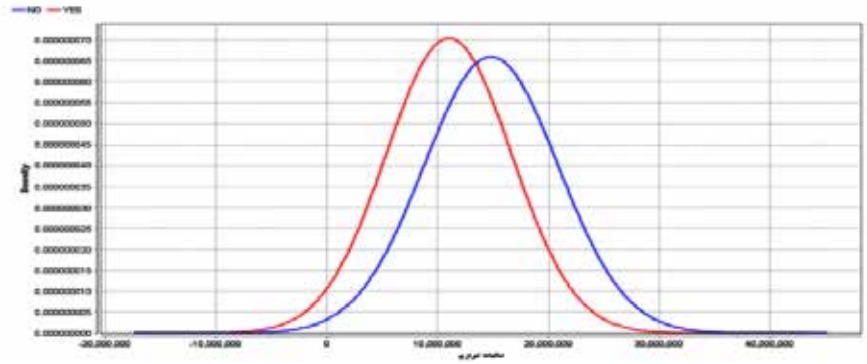
تحلیل نتایج

بر اساس خروجی داده کاوی پرونده‌های مالیاتی اصناف مطابق شکل زیر مشاهده می‌شود میانگین مالیات ابرازی اصنافی که فرار مالیاتی داشته‌اند از میانگین مالیات ابرازی اصنافی که فرار مالیاتی نداشته‌اند کمتر می‌باشد.

[Downloaded from taxjournal.ir on 2026-11-07]

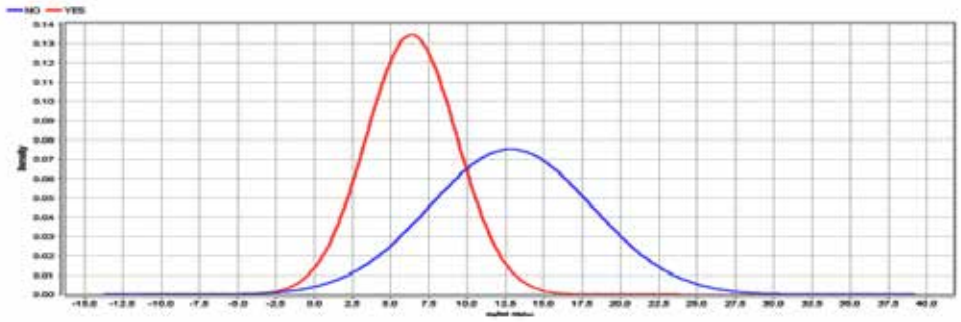
[DOI: 10.52547/taxjournal.30.56.1]

شکل (۲)ـ بررسی مالیات ابرازی اصناف دارای فرار مالیاتی



همچنین مشاهده می‌شود میانگین سابقه فعالیت اصنافی که فرار مالیاتی داشته‌اند(۶سال) از میانگین مالیات ابرازی اصنافی که فرار مالیاتی نداشته‌اند (۹سال) کمتر می‌باشد.

شکل (۳)ـ بررسی سابقه فعالیت اصناف دارای فرار مالیاتی



همچنین بر اساس خوشه‌بندی اصناف مورد مطالعه مشاهده می‌شود:

- در تمامی خوشه‌ها فرار مالیاتی وجود دارد ولی نسبت فرار مالیاتی به کل ۳۲/۵درصد از کل می‌باشد.
- بر اساس داده‌های مورد تحلیل مشاهده می‌شود، گروه صنفی صاحبان دفاتر رسمی و حقوقی و مشاغل وابسته(G1) فرار مالیاتی کمتری نسبت به سایر گروه‌های صنفی دیگر دارد. همچنین گروه صنفی نمایشگاه و فروشگاه لوازم خودرویی و مشاغل وابسته(G5) فرار مالیاتی بیشتری نسبت به سایر خوشه‌های صنفی دیگر را دارد.

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

- بیشتر نمونه‌های مورد تحلیل در گروه صنفی نمایشگاه و فروشگاه لوازم خودرویی و مشاغل وابسته G5 (۸۰۸ تعداد نمونه)می باشندو کمترین گروه صنفی صاحبان دفاتر رسمی و حقوقی و مشاغل وابسته G1 (۲۹۲ نمونه) می‌باشد.
- گروه صنفی صاحبان دفاتر رسمی و حقوقی و مشاغل وابسته (G1) فرار مالیاتی کمتری نسبت به سایر گروه‌های صنفی دیگر دارد همچنین گروه صنفی نمایشگاه و فروشگاه لوازم خودرویی و مشاغل وابسته (G5) فرار مالیاتی بیشتری نسبت به سایر خوشه‌های صنفی دیگر را دارد.
- در گروه صنفی صاحبان دفاتر رسمی و حقوقی و مشاغل وابسته (G1) بیشتر فرار مالیاتی مربوط به صنوفی می‌باشد که سابقه فعالتی کمتر ۱۱ سال داشته‌اند.

نتیجه‌گیری و پیشنهادات

در این پژوهش یک الگوی پیش‌بینی فرار مالیاتی اصناف مبتنی بر تکنیک‌های داده کاوی ارائه گردید، که بر اساس جستجوی انجام شده و خلاءهای پژوهش‌های داخلی و خارجی تا کنون پژوهشی در این زمینه انجام نشده بود و این کار برای اولین بار در حوزه امور مالیاتی اصناف استان قزوین اجرا گردید. این پژوهش شامل بررسی و تحلیل کلیه صنوف دارای کد مالیاتی در استان قزوین بین سال‌های ۹۸-۱۳۹۳ است. در این پژوهش پرونده مالیاتی مرتبط با اصناف در پنج گروه مالیاتی به عنوان داده‌های نمونه انتخاب گردید این پنج گروه عبارت است از: گروه صنفی صاحبان دفاتر رسمی و حقوقی، گروه صنفی مشاورین املاک، گروه صنفی تالارهای پذیرایی، رستوان و مشاغل وابسته، گروه صنفی خدمات ارتباطی و گروه صنفی نمایشگاه و فروشگاه لوازم خودرویی و مشاغل وابسته می‌باشند. شش مرحله در پیش پردازش داده در فرایند داده کاوی شامل یکپارچه‌سازی داده، تجمیع داده، نمونه‌گیری، پاک‌سازی داده‌ها، انتخاب زیر مجموعه‌ای از ویژگی‌ها و تغییر شکل متغیرها صورت پذیرفت. بر اساس تجزیه و تحلیل داده‌ها نتایج نشان می‌دهد، مدل درخت تصمیم بر اساس داده‌های موجود مدل مناسبی محسوب می‌شود و عملکرد مناسب جهت پیش‌بینی رفتار الگوی فرار مالیاتی را دارد. بر اساس ماتریس آشفتگی مدل ساز بر اساس دو کلاس فرار مالیاتی دارد، ۶۷۲ نمونه که در کلاس فرار مالیاتی ندارد (NO) را درست پیش‌بینی نمود. لذا ۶۰٪ پیش‌بینی در کلاس فرار مالیاتی ندارد (NO) را مدل ساز به درستی پیش‌بینی کرده است که نشان از صحت و دقت بالا در پیش‌بینی دارد. همچنین ۴۰۲ نمونه که در کلاس فرار مالیاتی دارد (YES) را درست پیش‌بینی کرده است. لذا ۸۳٪ درصد پیش‌بینی در کلاس فرار مالیاتی دارد (YES) را مدل ساز به درستی پیش‌بینی کرده است که نشان از صحت و دقت بالا در پیش‌بینی دارد.

نتایج این پژوهش در بخش تجزیه و تحلیل داده نشان می‌دهد الگوی پیش‌بینی فرار مالیاتی اصناف مبتنی

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

مدل درخت تصمیم عملکرد مناسب را دارد و نتایج این پژوهش اعتبار لازم جهت تصمیم‌گیری مدیران امور مالیاتی استان در مورد ایجاد مکانیزم اجرایی کنترل فرار مالیاتی پنج گروه اصناف مورد مطالعه را دارد. لذا به مدیران امور مالیاتی استان پیشنهاد می‌شود، از نتایج این پژوهش جهت تحلیل داده‌های مرتبط با اصناف با توجه به محدودیت منابع انسانی در دسترس استفاده نمایند. همچنین با توجه به وجود داده‌های مختلف در پایگاه‌های داده موجود نسبت به تحلیل‌های دوره‌ای بر اساس الگوریتم‌های داده کاوی بیش از ۳۲ درصد از اصنافی که مالیات ابرازی خود را اعلام نموده‌اند فرار مالیاتی داشته‌اند لذا به مدیران امور مالیاتی استان پیشنهاد می‌شود نسبت به تدوین برنامه‌های راهبردی کنترلی نسبت به کاهش فرار مالیاتی اصناف اقدام لازم را انجام دهند. نتایج داده کاوی نشان می‌دهد گروه صنفی نمایشگاه و فروشگاه لوازم خودرویی و مشاغل وابسته فرار مالیاتی بیشتری نسبت به سایر خوشه‌های صنفی دیگر دارند. همچنین نتایج نشان می‌دهد در گروه صنفی صاحبان دفاتر رسمی و حقوقی و مشاغل وابسته بیشتر فرار مالیاتی مربوط به صنوفی می‌باشد که سابقه فعالیت کمتر از ۱۱ سال دارند لذا پیشنهاد می‌گردد مأموران مالیاتی کنترل بیشتری نسبت به فرار مالیاتی این گروه از اصناف را داشته باشند. نتایج داده کاوی نشان می‌دهد میانگین سابقه فعالیت اصنافی که فرار مالیاتی داشته‌اند از میانگین مالیات ابرازی اصنافی که فرار مالیاتی نداشته‌اند کمتر می‌باشد. به مأموران مالیاتی توصیه می‌شود که کنترل بیشتری نسبت به فرار مالیاتی گروهی از اصناف با سابقه کمتر از ۶ سال را داشته باشند. جهت داده کاوی بررسی فرار مالیاتی اصناف از پایگاه داده امور مالیاتی استان قزوین به عنوان مورد مطالعه در یک سری زمانی مشخص استفاده شده است. به سایر پژوهشگران پیشنهاد می‌گردد با توجه به کلاس‌بندی منتخب در این پژوهش، داده‌های سایر استان‌های کشور را مورد تجزیه و تحلیل قرار داده و نتایج را جهت بررسی جامعیت نتایج مقایسه نمایند. در این پژوهش به دلیل اهمیت موضوع و محدودیت‌های زمانی، داده‌های مالیاتی مرتبط با اصناف در پنج گروه مالیاتی شامل گروه صنفی صاحبان دفاتر رسمی و حقوقی، گروه صنفی مشاورین املاک، گروه صنفی تالارهای پذیرایی، رستوان و مشاغل وابسته، گروه صنفی خدمات ارتباطی و گروه صنفی نمایشگاه و فروشگاه لوازم خودرویی و مشاغل وابسته مورد تجزیه و تحلیل قرار گرفت. در پژوهش‌های بعدی توصیه می‌شود که سایر گروه‌های اصنافی که در این پژوهش به آنها پرداخته نشده است انتخاب و مورد مطالعه قرار گیرد. از مهم‌ترین محدودیت‌های این پژوهش می‌توان به عدم دسترسی کامل به اطلاعات گروه صنفی صاحبان دفاتر رسمی و حقوقی و مشاغل وابسته در محدوده زمانی مورد نظر اشاره نمود. همچنین به دلیل شرایط خاص دوران کرونا، دسترسی به طیف وسیعی از خبرگان حوزه مالیاتی دشوار بوده و محقق برای دریافت نظرات تیم خبرگان در این پژوهش، از روشهای مجازی استفاده نموده است.

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

فهرست منابع

1. Rezaqoli Zadeh M, Aghii M, Alami A.H. (2019). Analysis of Tax Evasion in Iran by Multiple Index_multiple Causes (MIMIC) Method. Majlis and Strategy. 26(97): 191 – 226, (Persian).
2. Khodaparast M, Sultan Hosseini M, Salimi M. (2019). Estimating the Relative Share of Managerial Factors on Tax Evasion of Athletes and Professional Coaches of Sports Clubs in Isfahan Province. Physiology and Management Researches in Sports. 10(2):101_113, (Persian).
3. Dehghani S, Mousavi Jahromi Ye, Abdoli Q. (2019). A New Approach in Explaining the Phenomenon of Tax Evasion. Economic Research. 53(1): 1 _ 23, (Persian).
4. Mohammadi Khiare M. (2019). Investigation of the Impact of Tax Evasion and Corruption on Entrepreneurship: the Case Study of OECD Countries. Entrepreneurship Development. 11(3): 501 – 520, (Persian).
5. Omidpour R, Pajhoyan J. (2018). Tax Evasion in the Income Tax Base of Legal Entities in Iran (Annual Estimates 1352_1392). Financial Economy (Financial Economy and Development). 11(39): 27 _ 56, (Persian).
6. Rezaei Siabidi M. (2018). Ways to Fight and Prevent Tax Evasion. Yar Law, 4(4): 145_ 159, (Persian).
7. Amiri M. (2018). Behavioral Economics and Tax Evasion. Economic Research Journal. 17(64): 95_ 130, (Persian).
8. Fetras M H, Delai Milan A. (2017). Investigating the Underground Economy and Tax Evasion in the Framework of Dynamic Stochastic General Equilibrium (DSGE) Models. Economic Growth and Development Studies 7(25): 65 _84, (Persian).
9. Karamkhani J, Weysmoradi A, Ali Madd Z. (2017). Investigating the Effectiveness of Tax Crimes in Preventing Tax Evasion in the Value Added Tax System among Taxpayers in Ilam Province. State Accounting. 2(4):25 _36, (Persian).
10. Karimi Patanlar S, Gilak Hakimabadi M T, Saber N F. (2016). Examining the Effect of Government Effectiveness on Reducing Tax Evasion in Selected Countries. Tax Research Journal. 23(27): 63 _ 90, (Persian).
11. Hamidi N, Mohammadzadeh A, Mohammadi F. (2016). Investigating the Position of Tax Crimes in Preventing Evasion in the Value Added Tax System (Case Study

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

of Qazvin Province). Tax Research. 23(27): 147 ـ 166, (Persian).

12. Maddah M, Khaleq Panah Z. (2016). Tax Evasion in Iran's Imports, the Approach of the Hybrid Model of Artificial Neural Network and Gradual Cooling Algorithm. Planning and Budgeting. 20(2): 85 ـ 102, (Persian).
13. Zahi N, Mohammad Khanali S. (2011). Investigation of Factors Affecting Tax Evasion (Case Study of East Azarbaijan Province). Tax Research Paper. 18(9): 25 ـ60, (Persian).
14. Taqvi Fard M.T, Raisi Vanani I, Panahi Re. (2018). Prospective Analysis of Tax Evasion Detection in Modian Value Added Tax using Classification and Clustering Algorithms. Tax Research Journal. 25(35): 11 ـ 35, (Persian).
15. Laridasht M, Qaim Masaki K, Kahrami Q. (2017). Investigation of Factors Affecting Tax Evasion in South Khorasan Province with Emphasis on Cultural Components. Value and Behavioral Accounting. 1(2): 139 164, (Persian).
16. Rahimi KIA E, Mohammadi S, Ghazanfari M. (2016). Tax Evasion Detection using a Combined Intelligent System. Tax Research Journal. 23(2): 135 ـ 163, (Persian).
17. Tasgir M, Gharibi M. (2016). Application of Data Mining Methods to Improve the Performance of Tax Evasion Detection. Tax Research Journal. 23(28): 95 ـ 116, (Persian).
18. Khosravi T, Pezhuyan J. (2014). The Effect of Corporate Taxes on Private Sector Investment using the Banks Approach, Financial Economics (Financial Economics and Development). 7(25): 95 ـ 121, (Persian).
19. Amiri R. (2019). Tax Avoidance, Tax Risk and Debt Maturity in Companies Listed on the Tehran Stock Exchange, Islamic Azad University, Marvdasht Branch, Master's Thesis, (Persian).
20. Panahi Re. (2017). Detection of Tax Evasion of VAT Taxpayers using Data Mining Methods (Case Study: VAT Taxpayers of Tehran), Allameh Tabatabai University, Master's Thesis, (Persian).
21. Moghimi Nia A. (2009). Designing a Suitable Solution for Determining and Calculating Business Tax Coefficients, Research Journal on Taxation. Scientific Quarterly Journal of the Tax Administration of Iran. 16 (3): 38_9, (Persian).
22. Arab Mazar A.A, Dehghani A. (2009). Estimating the Efficiency of Taxes on the Income of Businesses and Legal Entities, Research Journal of Taxation. 17(7):

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

45 ـ 64, (Persian).
23. Khosro P, Elizabeth Soltani S. (2009). Estimation of Tax Capacity in Fars Province, Tax Research Journal, Scientific Quarterly of the National Tax Affairs Organization, 16(1): 29_49, (Persian).
24. Yamen A. (2018). Impact of Institutional Environment Quality on Tax Evasion: A Comparative Investigation of Old Versus New EU Members. Journal of International Accounting, Auditing and Taxation. 32: 17_29.
25. Abdixhiku L. B, Krasniqi G. (2017).Pugh and I. Hashi, Firm_level Determinants of Tax Evasion in Transition Economies, Economic Systems.41(3):354_366.
26. Stankevicius E. (2015). Hybrid Approach Model for Prevention of Tax Evasion and Fraud. Social and Behavioral Sciences.213: 383_389.
27. Khalil S and Sidani Y. (2020). The Influence of Religiosity onTax Evasion Attitudes in Lebanon. Journal of International Accounting. Auditing and Taxation. 40: 220_235.
28. Agarwal S, Li K, Yu Qin, Jing W, Yan J. (2020). Tax Evasion, Capital Gains Taxes, and the Housing Market. Journal of Public Economics. 188: 104_122.
29. Di Gioacchino D, Fichera D. (2020). Tax Evasion and Tax Morale: A Social Network Analysis. European Journal of Political Economy. 65: 121_149.
30. Demir B, Javorcik B. (2020).Trade Policy Changes, Tax Evasion and Benford's Law. Journal of Development Economics. 144: 152_178.
31. Yousefi K, Vesal M, Pilvar Hanifa. (2020). Import Tax Evasion and Avoidance: Evidence from Iran.The Quarterly Review of Economics and Finance. 75: 31_39.
32. Ruan J, Yan Z, Dong B, Zheng Q, Qian B. (2019). Identifying Suspicious Groups of Affiliated_transaction_based Tax Evasion in Big Data. Information Sciences. 477: 508_532.
33. Hung F.S. (2017). Explaining the Nonlinearity of Inflation and Economic Growth: The Role of Tax Evasion .International Review of Economics & Finance. 52: 436_445.
34. Didimo W, Giamminonni L, Liotta G, Montecchiani F, Pagliuca D. A. (2018). Visual Analytics System to Support Tax Evasion Discovery. Decision Support Systems. 110: 71_83.
35. González P.C, Velásquez J.D. (2017). Characterization and Detection of Taxpayers with False Invoices using Data Mining Techniques. Expert Syst. Appl. 40 (5): 60_81.

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]

36. Assylbekov Z, Melnykov I, Bekishev R, Baltabayeva A, Bissengaliyeva D, Mamlin E. (2016). Detecting Value_added Tax Evasion by Business Entities of Kazakhstan. Intelligent Decision Technologies Springer. 56: 37–49.
37. Chen Y. S, Cheng C. H. (2010). A Delphi_based Rough Sets Fusion Model for Extracting Payment Rules of Vehicle License Tax in the Government Sector. Expert Syst. Appl. 37(3): 2161–2174.
38. Devereux M.P, Griffith R, Klemm A. (2002). Can International Tax Competition Explain Corporate Income Tax Reforms. Econ. 35: 323_341.
39. Ferrantino M.J, Liu X, Wang Z. (2012). Evasion Behaviors of Exporters and Importers: Evidence from the US–China Trade Data Discrepancy. J. Int. Econ. 86 (1): 141–157.
40. Goumagias N.D, Hristu_Varsakelis D, Saraidaris A. (2012). A Decision Support Model for Tax Revenue Collection in Greece. Decis. Support Syst. 53 (1): 76–96.
41. Hsu K. W, Pathak N, Srivastava J, Tschida G, Bjorklund E. (2015). Data Mining Based on Tax Audit Selection: a Case Study of a Pilot Project at the Minnesota Department of Revenue. In: Real World Data Mining Applications, Springer. 89: 221–245.
42. Jia S. (2016). Tax Losses in China: Estimates, Analysis and Countermeasures. Friends Account. 13: 39–40.
43. Sikka P. (2010). Corporate Social Responsibility and Tax Avoidance in Accounting Forum. 34: 153–168.
44. Tian F, Lan T, Chao K.M, Godwin N, Zheng Q, Shah N, Zhang F.(2016). Mining Suspicious Tax Evasion Groups in Big Data. IEEE Trans, Knowl. Data Eng. 28 (10): 2651–2664.

[Downloaded from taxjournal.ir on 2026-11-07]

[DOI: 10.52547/taxjournal.30.56.1]