

ارزیابی اثربخشی الگوریتم درخت تصمیم و رگرسیون چند متغیره خطی در پیش‌بینی فرار مالیاتی مؤذیان حقوقی استان مازندران

محسن مقری گردوباری

دانشجوی دکتری حسابداری، واحد ساری، دانشگاه آزاد اسلامی، ساری، ایران

mohsen_moghri1366@yahoo.com

ایمان داداشی

استادیار گروه حسابداری، دانشگاه قم، ایران (نویسنده مسئول)

I.dadashi@qom.ac.ir

بهرام محسنی ملکی

استادیار گروه حسابداری، دانشگاه مازندران، مازندران، ایران

b.mohsenim@umz.ac.ir

علی ذبیحی

استادیار گروه حسابداری، واحد ساری، دانشگاه آزاد اسلامی، ساری، ایران

azabihii.95@yahoo.com

درآمدهای مالیاتی یکی از مهم‌ترین منابع درآمدی دولت و تأمین کننده بخش عمده‌ای از هزینه‌های آن می‌باشد. هدف اصلی پژوهش حاضر، ارزیابی توان پیش‌بینی تکنیک‌های الگوریتم درخت تصمیم و رگرسیون چند متغیره خطی در پیش‌بینی فرار مالیاتی مؤذیان حقوقی بوده است. براساس مبانی نظری و پیشنهاد پژوهش‌های صورت گرفته در این حوزه، مجموعه‌ای مشتمل از ۵۷ شاخص مالی و غیرمالی در سه سطح کلان اقتصادی، مؤذیان و حسابرسان مالیاتی، در نمونه‌ای مشتمل بر ۹۶۴ پرونده مالیاتی اشخاص حقوقی اداره کل امور مالیاتی مازندران برای سال‌های ۱۳۹۸-۱۳۹۷ با استفاده از نرم افزار متلب و استاتا مورد بررسی قرار گرفته است. در این پژوهش، پس از انتخاب متغیرهای اثرگذار با استفاده از الگوریتم شناسایی سینوس-کسینوس، اقدام به پیش‌بینی فرار مالیاتی با بهره گیری از تکنیک‌های الگوریتم درخت تصمیم و رگرسیون شده است. نتایج حاصل از بررسی داده‌ها نشان داد که متغیرهای سطح مؤذیان و حسابرسان مالیاتی جهت پیش‌بینی فرار مالیاتی اثربخشی بیشتری دارند. همچنین یافته‌ها حاکی از آن بوده که توان پیش‌بینی الگوریتم درخت تصمیم نسبت به رگرسیون چند متغیره خطی، بیشتر است.

H26: JEL طبقه‌بندی

واژگان کلیدی: فرار مالیاتی، الگوریتم سینوس کسینوس، الگوریتم درخت تصمیم، رگرسیون.

۱. مقدمه

یکی از مظاهر تجلی استقلال کشورها، توانایی آنها در تأمین اعتبار برای توسعه با استفاده از منابع مالی ناشی از تولید ناخالص داخلی است (دجاجان تی^۱، ۲۰۲۰). امروزه مالیات، اساسی‌ترین منبع درآمدی دولت و مهمترین ابزار در تأمین و تخصیص بهینه منابع و افزایش میزان سرمایه گذاری در تولید ناخالص ملی می‌باشد (آنان^۲ و همکاران، ۲۰۱۳).

مالیات مهم‌ترین 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 (یامن^۳، ۲۰۲۱). همواره تلاش برای بهره گیری 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 (آبدیخیکو^۴ و همکاران، ۲۰۱۹). در حال حاضر حسابرسان مالیاتی با چالش شناسایی و جمع‌آوری مالیات از افراد روبرو می‌باشد که به طور موفقیت‌آمیزی از پرداخت مالیات صحیح فرار می‌کنند. به منظور رویارویی با این چالش حسابرسان مالیاتی به منابع محدود و راهبردهای حسابرسی سنتی تجهیز شده‌اند که زمان بر می‌باشد (آگاروال^۵، ۲۰۲۰). 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 شیوه‌هایی که بتوانند فرار مالیاتی را شناسایی و کشف کنند، مسئله مهم و چالش برانگیز بوده است. اگر دولت نتواند فرار مالیاتی را کشف کند، سرمایه‌گذاری‌های عمومی به دلیل کاهش مالیات‌ها و در پی آن کاهش درآمدهای دولت به طور منفی تحت تأثیر قرار خواهد گرفت (گیوچینو^۶، ۲۰۲۰).

-
1. Djajanti
 2. Annan
 3. Yamen
 4. Abdixhiku
 5. Agarwal
 6. Di Gioacchino

براساس برنامه پنجم توسعه، اقتصادی، اجتماعی و فرهنگی جمهوری اسلامی ایران، دولت مکلف به قطع وابستگی هزینه‌های جاری دولت به درآمدهای نفت و گاز تا پایان برنامه شد و با عدم تحقق آن، مجدداً در ماده ۶ قانون برنامه ششم نیز تکرار شد. این در حالی است که برابر تحقیقات صورت گرفته میزان فار مالیاتی در ایران به طور میانگین ۲۲/۳۷ درصد تولید ناخالص داخلی (صمدی و تابنده، ۱۳۹۲؛ صادقی و شکیابی، ۱۳۸۰؛ میلانی و اکبرپور، ۱۳۹۱)، نسبت آن به هزینه‌های جاری دولت تا پایان سال ۱۳۹۱ ۴۳/۷ درصد و سهم آن در بودجه عمومی کشور ۳۶/۵ درصد است (عسکری، ۱۳۹۲). تحقیقات نشان می‌دهد که میزان فار مالیاتی در ایران در بازه زمانی سال ۱۳۵۲ تا ۱۳۹۲ معادل ۵۴ درصد کل تولید ناخالص داخلی کشور بوده است.

یکی از پررنگ‌ترین بخش منابع قانون بودجه ۱۴۰۲ موضوع مالیات‌هاست. جزئیات منابع بودجه دولت نشان می‌دهد که پیش‌بینی درآمدهای مالیاتی حدود ۵۴ درصد رشد نسبت به سال قبل خواهد داشت. درآمدهای مالیاتی مطابق قانون بودجه ۱۴۰۲ سهم درآمدهای دولت به شرح جدول

(۱) پیش‌بینی گردید:

جدول ۱. درآمدها (منابع) بودجه سال ۱۴۰۲ کشور

سرفصل	مبلغ (به میلیون ریال)	درصد سهم
درآمدهای مالیاتی (به غیر از مالیات واردات)	۷.۴۸۹.۶۰۱.۶۷۵	% ۲۱/۱
مالیات بروارادات (گمرک)	۱.۴۵۴.۹۳۰.۰۰۰	% ۱۳/۸
درآمد حاصل از مالکیت دولت	۱.۰۴۵.۴۲۹.۰۰۰	% ۹/۹
درآمد حاصل از فروش کالا و خدمات	۲۶۰.۰۷۱.۱۰۰	% ۲/۵
درآمد حاصل از جرائم و خسارت	۸۰.۳۶۱.۰۰۰	% ۰/۸
درآمد متفرقه	۲۱۰.۴۴۵.۸۹۹	% ۰/۲
جمع	۱۰.۵۴۰.۸۳۸.۶۷۴	% ۱۰۰

مأخذ: قانون بودجه ۱۴۰۲

نتایج برخی تحقیقات حاکی از آن است که میانگین نسبت فار مالیاتی برآورده به کل درآمدهای مالیاتی و تولید ناخالص داخلی به ترتیب ۷۳/۵ و ۵۴/۶ درصد بوده (امیدی پور و دیگران، ۱۳۹۴) که بیانگر حجم بالای فار مالیاتی در اقتصاد ایران و نظام مالیاتی کشور می‌باشد. بر این اساس

با توجه به اینکه داده‌های این پژوهش که از مؤدیان اشخاص حقوقی اداره کل امور مالیاتی استان مازندران احصاء شده در جدول شماره (۲) اختلاف بین مالیات ابرازی در اظهارنامه و مالیات قطعی برای تمامی مؤدیان اشخاص حقوقی اداره کل امور مالیاتی مازندران ارائه گردید.

جدول ۲. اختلاف بین مالیات ابرازی و مالیات قطعی مؤدیان اشخاص حقوقی مازندران (مبالغ به هزار ریال)

سال	تعداد اظهارنامه دریافتی	مبلغ مالیات ابرازی در اظهارنامه مؤدیان	مبلغ مالیات در برگ قطعی	نسبت مالیات ابرازی به مالیات قطعی
۱۳۹۲	۹۱۲۶	۱.۰۳۹.۸۵۷.۸۱۵	۲.۷۳۳.۷۶۸.۶۴۵	% ۳۸
۱۳۹۳	۸۴۳۲	۸۶۹.۱۷۸.۴۳۹	۲.۸۰۴.۵۶۰.۹۰۷	% ۳۱
۱۳۹۴	۹۵۴۵	۱.۲۲۸.۷۰۳.۱۹۷	۲.۴۰۹.۳۷۱.۷۱۱	% ۵۱
۱۳۹۵	۸۸۵۱	۹۵۷.۹۱۶.۰۲۷	۲.۹۸۵.۰۵۵.۶۸۱	% ۳۲
۱۳۹۶	۸۹۷۱	۷۴۷.۷۲۰.۴۵۹	۲.۵۲۶.۷۶۱.۰۱۸	% ۳۰
۱۳۹۷	۱۱۷۱۴	۶۲۷.۲۴۸.۱۳۳	۳.۸۸۶.۱۸۷.۷۲۱	% ۱۶
۱۳۹۸	۱۱۳۰۷	۲.۳۹۲.۶۵۱۵۰۳	۵.۸۹۸.۵۱۸.۷۶۰	% ۴۰
۱۳۹۹	۱۱۲۵۱	۷.۷۶۷.۴۴۴.۶۴۴	۱۲.۶۹۱.۹۷۷.۲۷۱	% ۶۱

منبع : یافته‌های محقق

با وجود فرار مالیاتی گسترده در نظام‌های مالیاتی، پژوهش‌های مرتبه با فرار مالیاتی در چند سال اخیر افزایش یافته است. دلایل فرار مالیاتی در کشورهای مختلف ناشی از تنوع پایه‌های مالیاتی و ترکیب نامناسب درآمدهای مالیاتی، وجود معافیت‌های مالیاتی وسیع و متنوع، تعدد، پیچیدگی و ابهام در قوانین مقررات و فرآیندهای مالیاتی، افزایش فساد و گسترده‌گی بخش غیر رسمی اقتصاد، عدم اعتماد مردم به کارایی و اثربخشی دولت، فرهنگ مالیاتی ضعیف، فقدان نظام جامع اطلاعاتی مالیاتی و مشکلات ساختار اقتصادی می‌باشد.

لذا، با توجه به وجود فرار مالیاتی در کشور در این پژوهش به ارائه الگویی جهت پیش‌بینی فرار مالیاتی با تأکید بر شاخص‌های سطح کلان اقتصادی، سطح مؤدیان مالیاتی و سطح حسابرسان مالیاتی اعم از متغیرهای مالی و غیرمالی پرداخته می‌شود؛ بنابراین با توجه به مطالب فوق، در این پژوهش ابتدا

متغیرهای اثرگذار بر فرار مالیاتی از طریق الگوریتم بهینه‌ساز سینوس کسینوس شناسایی و سپس از میان الگوهای هوش مصنوعی و آماری، الگوریتم درخت تصمیم و رگرسیون چند متغیره خطی با یکدیگر مقایسه تا مشخص شود کدام یک از الگوهای ذکر شده توان بیشتری برای پیش‌بینی فرار مالیاتی مؤیدیان اشخاص حقوقی اداره کل امور مالیاتی مازندران دارد تا ماموران مالیاتی در کشور به خوبی بتوانند وجود فرار مالیاتی در مؤیدیان حقوقی، رسیدگی و درآمد مشمول مالیات را تشخیص دهند.

ساختار مقاله حاضر بدین صورت است که در بخش اول به مبانی نظری فرار مالیاتی و شاخص‌های آن پرداخته می‌شود. در بخش دوم، پیشینه مطالعات تجربی مورد بررسی قرار می‌گیرد. در ادامه، به توصیف روش پژوهش و متغیرهای پژوهش می‌پردازد. سپس، روش‌های تجزیه و تحلیل داده‌ها و یافته‌های پژوهش مطرح می‌شود. سرانجام، نتیجه گیری، پیشنهادها و محدودیت‌های پژوهش بیان می‌گردد.

۲. مبانی نظری

رویکرد مالیاتی مؤیدیان در قالب طرح ریزی مالیات در طیفی از تمکین تا عدم تمکین مالیاتی فرار دارد. عدم تمکین مالیاتی خود شامل دامنه‌ای وسیع از اجتناب مالیاتی، تهور مالیاتی، سپرسازی مالیاتی و فرار مالیاتی است (اعتمادی و همکاران، ۱۳۹۵). عدم تمکین یعنی مؤدی به هر دلیلی در تکمیل اظهارنامه مالیاتی، گزارش بدھی مالیاتی، محاسبه و پرداخت دقیق آن کوتاهی کند (برزیده و ابراهیمی درده، ۱۳۹۸). فرار از پرداخت مالیات در واقع عملی عمدی با هدف عدم تمکین قانون مالیاتی به منظور کاهش بدھی مالیاتی است (محبوبی و شهبازی، ۱۳۹۰). این امر به شیوه‌های مختلفی مانند کم اظهاری درآمد مشمول مالیات، بیش اظهاری هزینه‌ها و زیان‌ها، عدم افشاری عمدی حقایق با اهمیت، عدم پرداخت مالیات و... انجام می‌گیرد (ناوانیتا و دارشانا، ۲۰۲۰).

هر گونه تلاش غیرقانونی برای پرداخت نکردن مالیات مانند ندادن اطلاعات لازم در مورد عواید و منافع مشمول مالیات به مقامات مسئول فرار مالیاتی نامیده می‌شود (یوسفی، ۱۳۹۹). در فرار مالیاتی به رغم تطبیق فعالیت با قوانین در اجرای فعالیت یک یا چند ضابطه مصوب مراجع ذی ربط تماماً

1. Nananeetha and Darshana

نادیده گرفته می‌شود (پناهی، ۱۳۹۸). وقتی مؤدی از گزارش درآمد مشمول مالیات خویش عدول کند یا به شیوه‌های عامدانه نظری گزارش کمتر از واقع درآمد و فروش، گزارش بیش از واقع کسورات و یا تنظیم اظهارنامه غیر واقعی 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 نفریق بالا تعریف می‌کند (روان، ۲۰۱۹).

به گونه کلی، فرار مالیاتی به عدم پرداخت غیرقانونی تمام یا قسمتی از مالیات اطلاق می‌شود (موسی، طهماسبی و خاکی، ۱۳۸۸). این اقدام می‌تواند در غالب تنظیم حساب‌های مالی، ارائه اطلاعات مالی نادرست و یا ارائه نکردن اطلاعات صورت گیرد. افزون بر این، فرار مالیاتی به گستره اقتصاد غیررسمی، مانند معاملات خانگی و فعالیت‌های غیرقانونی نظری فعالیت‌های جنابی و بزهکاری مربوط می‌شود که از این فعالیت‌ها به دلیل فقدان اطلاعات لازم، مالیاتی اخذ نمی‌گردد (زهی و همکاران، ۱۳۸۹). فرار مالیاتی را می‌توان یکی از مظاهر فساد اقتصادی دانست. از این رو، تمامی کشورها تلاش خود را برای کاهش این پدیده به کار می‌گیرند و در این راستا همواره به اصلاح نظام مالیاتی خود می‌پردازنند. فرار از مالیات چه به صورت مستقیم و چه به صورت غیر مستقیم بر فعالیت‌های اقتصادی از جمله سرمایه‌گذاری مؤثر بوده و می‌تواند در بلند مدت مسیر رشد و توسعه اقتصادی را دچار اختلال کند (شهریاری فر، ۱۳۹۰).

مروری بر ادبیات موضوع، میین آن است که اساساً دو جریان عمدۀ در تبیین رفتار فرار مالیاتی وجود داشته است. از یک سوء مدل مبتنی بر اقتصاد نئوکلاسیک مرسوم که سعی بر آن داشته‌اند تا پدیده تمکین یا عدم تمکین مالیاتی را بر بنیان داده‌های مفروض در اقتصاد متعارف، از قبیل اصل عقلانیت فردی و اصل حد اکثرسازی سود مورد انتظار تبیین کنند و از سوی دیگر، رویکرد متفاوتی را شاهد هستیم که یا به کلی، منتقد مدل فرمالیستی^۱ اقتصاددانان مرسوم بوده‌اند یا پیش فرض‌های

1. Ruan

۲. فرمالیسم همان‌گونه که از نامش بر می‌آید، بر فرم تأکید دارد تا محتوا. آنچه تحت عنوان فرمالیسم در علم اقتصاد، در محیط‌های آکادمیک ایران و نیز کشورها از آن صحبت می‌شود، کاربرد بیش از حد ریاضیات و ابزارهای اقتصاد سنجی در آموزش و پژوهش‌های اقتصادی است؛ روندی که به جریان غالب یا متعارف علم اقتصاد بدل شده است.

اولیه آنها را در تبیین این پدیده کافی نمی‌دانند و در صدد افزودن پارامترهای غیراقتصادی، مانند نهادهای رسمی و غیررسمی به مدل‌های ریاضی وار آنها برآمده‌اند (بیکر، ۱۹۶۸). نمونه نوعی جریان نخست، مدل پیشنهادی آلینگهام و ساندمو (۱۹۷۲) (یا همان الگوی A-S) است. مرلی (۱۹۸۴)، به نقل از میلانی و اکبرپور، (۱۳۹۱) به تعمیم مدل مذکور می‌پردازد تا بتواند رفتار تمکین عدم تمکین یا اشخاص حقوقی را تبیین کند. پژوهشگران دیگری مانند چن و چو (۲۰۰۲) (به نقل از میلانی و اکبرپور، ۱۳۹۱)، نیز مدل استاندارد را از نقطه نظر مالیات بر اشخاص حقوقی مورد ارزیابی قرار داده‌اند. براساس مطالعه کرچلر و همکاران (۲۰۱۴)، رویکرد تمکین مالیاتی شخص مبتنی بر نظریه شب لغزنده یا رفتار جبرانی شکل می‌گیرد که منشاء آن دو بعد اعتماد مؤدی و قدرت سازمان مالیاتی است. دجاجان تی (۲۰۲۰) بعد سومی را نیز به صورت منصفانه بودن نظام مالیاتی به ابعاد فوق اضافه می‌نماید. البته این پژوهشگران قائل به پویایی رویکرد و رفتار مالیاتی مؤدیان بودند.

بر اساس واکاوی قانون مالیات‌های مستقیم، طبق ماده ۱۹۴ قانون فوق، مؤدیان مالیاتی که اظهارنامه مالیاتی آنها در اجرای مقررات ماده ۱۵۸ این قانون مورد رسیدگی قرار می‌گیرد در صورتی که درآمد مشمول مالیات مشخصه قطعی با رقم اظهارشده از طرف مؤدی مالیاتی بیش از ۱۵ درصد اختلاف داشته باشد علاوه بر تعلق جرائم مقرر مربوط که قابل بخشودن نیز نخواهد بود تا سه سال بعد از ابلاغ مالیات مشخصه از هر گونه تسهیلات و بخشودگی‌های مقرر در قانون مالیات‌ها نیز محروم خواهند شد. همچنین بر اساس ماده ۲۷۴ قانون مالیات‌های مستقیم جزو موارد جرم مالیاتی محسوب شده و به مجازات حبس تعزیری درجه شش^۱ محکوم می‌گردد (قانون مالیات‌های مستقیم، ۱۳۹۴).

۱. بر اساس ماده ۱۹ قانون مجازات اسلامی مصوب ۱۳۹۲/۰۲/۰۱، مجازات‌های تعزیری به هشت درجه تقسیم می‌شود. درجه شش آن عبارتست از: «حبس بیش از شش ماه تا دو سال»، «جزای نقدی بیش از بیست میلیون ریال تا هشتاد میلیون ریال»، «شلاق از سی و یک تا هفたاد و چهار ضریبه»، «محرومیت از حقوق اجتماعی بیش از شش ماه تا پنج سال»، «انتشار حکم قطعی در رسانه‌ها»، «ممنوعیت از یک یا چند فعالیت شغلی یا اجتماعی برای اشخاص حقوقی حداقل تا مدت پنج سال»، «ممنوعیت از دعوت عمومی برای افزایش سرمایه برای اشخاص حقوقی تا مدت پنج سال».

۳. پیشینه پژوهش

۱-۳. مطالعات داخلی

صیام و همکاران (۱۴۰۲) در مقاله‌ای تحت عنوان بررسی عوامل مؤثر بر فرار مالیاتی در شرکت‌های سخت مالیات ده با استفاده از روش اثر گذاری غیرخطی پرداختند. در بخش نخست به بررسی عوامل فرار مالیاتی با استفاده از روش رگرسیون غیرخطی طی دوره زمانی ۱۴۰۰-۱۳۷۰ پرداخته شده است. متغیرهای نرخ مؤثر مالیاتی شرکت، اندازه دولت، ابهام قوانین و مقررات و تعداد پرونده‌ها در چرخه دادرسی مالیاتی به عنوان متغیرهای اثرگذار در نظر گرفته شد. نتایج آزمون غیرخطی بودن الگو نشان می‌دهد نرخ مؤثر مالیاتی دارای رابطه غیرخطی با متغیر فرار مالیاتی می‌باشد.

ربیع بیگی و همکاران (۱۴۰۲) در پژوهشی تحت عنوان ارائه الگوی روحیه مالیاتی با رویکرد تئوری زمینه بنیان پرداختند. رویکرد این پژوهش ترکیبی از روش‌های کمی و کیفی است. در بخش کیفی مبتنی بر نظریه داده بنیاد مدل خود را ارائه می‌دهد و در بخش کمی از مدل سازی معادلات ساختاری استفاده شده است. جامعه آماری در بخش کیفی ۱۲ نفر از خبرگان مالیاتی و دانشگاهی و در بخش کمی ۲۷۰ نفر از مدیران مالی و کارشناسان مالیاتی بوده است. نتایج نشان می‌دهد که الگوی روحیه مالیاتی حول مقوله محوری فضای مجازی، دموکراسی و تبلیغات قرار دارد و منجر به تقویت روحیه مالیاتی می‌شود.

نجفی و همکاران (۱۴۰۱) در پژوهشی با ارائه الگویی برای کشف تقلب مالیاتی بر اساس تیپ‌های شخصیتی مدیران مالی با استفاده از رویکرد شبکه‌های عصبی پرداختند. جامعه آماری پژوهش را کلیه مدیران مالی شرکت‌های بورسی و غیربورسی در سال ۱۳۹۹ که معاف از مالیات نیستند، تشکیل می‌دهند. نتایج پژوهش حاکی از آن است که الگوی شبکه عصبی طراحی شده با ۱۰ نرون در لایه پنهان دارای دقت ۷۹/۵ درصد توانایی کشف تقلب مالیاتی صورت گرفته توسط مدیران مالی را دارد. همچنین طبق نتایج حاصل از مدل رگرسیونی، تیپ‌های شخصیتی روان رنجوری، برون‌گرایی، انعطاف‌پذیری و توافق‌پذیری بر تقلب مالیاتی تأثیر مثبت و معناداری و تیپ شخصیتی باوجود آن بودن تأثیر منفی و معنی داری بر تقلب مالیاتی دارد.

قاسمی و همکاران (۱۴۰۱) در مقاله‌ای تحت عنوان ارائه الگوی پیش‌بینی فرار مالیاتی اصناف مبتنی بر تکنیک داده کاوی پرداختند. داده‌های مورد تحلیل شامل بررسی ۵۶۰۰ پرونده مالیاتی کلیه اصناف دارای کد مالیاتی در استان قزوین طی سال‌های ۱۳۹۳ تا ۱۳۹۸ می‌باشد. برای مدل سازی در این مقاله از الگوی کلاس بندی درخت تصمیم استفاده گردید. نتایج نشان می‌دهد، مدل درخت تصمیم بر اساس داده‌های موجود، مدل مناسبی جهت پیش‌بینی محسوب می‌شود. معیار پوشش برابر با ۶۸٪ به دست آمده است که عملکرد خوب مدل‌ساز را نشان می‌دهد.

فلاح شمس و همکاران (۱۴۰۱) در مقاله مقایسه مدل شبکه عصبی و رگرسیون لاجیت در اعتبارسنجی مؤدیان مالیاتی طی سال‌های ۹۲ تا ۹۷ برای شرکت‌های فعال در بورس تهران پرداختند. در این پژوهش ابتدا با استفاده از تکنیک دلفی و متا سنتز ۱۶۴ مؤلفه مؤثر در اعتبارسنجی مؤدیان مالیاتی شناسایی گردید. یافته‌ها نشان می‌دهد که هر چند مدل لاجیت در شناسایی مؤدیان پررسیک توان بیشتری دارد، ولی در سایر موارد تشخیص مناسبی ندارد و در مجموع بر اساس نتایج حاصل از مقدار^۱ AUC مدل شبکه عصبی برازش بهتری دارد.

سل موسوی و همکاران (۱۳۹۹) به ارائه مدل پیش‌بینی فرار مالیاتی بر مبنای الگوریتم درخت تصمیم ID3 و شبکه بیزین مبتنی بر نسبت‌های مالی پرداختند. نتایج حاصل نشان داد که الگوریتم پیشنهادی با ۵۸/۶۰ درصد دقت، دارای بالاترین دقت صحت و با ۴۳/۷۶ درصد اشتباه دارای کمترین میزان اشتباه می‌باشد که الگوریتم‌های ID3 و الگوریتم بیزین که الگوریتم‌های پایه روش پیشنهادی می‌باشند، نیز بسیار بهتر عمل می‌کند.

مسیحی و همکاران (۱۳۹۸) در مقاله‌ای تحت عنوان استفاده از تکنیک‌های داده کاوی به سنجش ریسک مالیاتی مؤدیان مالیات بر ارزش افزوده پرداختند. جامعه آماری مقاله شامل اشخاص حقوقی بزرگ در شهر تهران می‌باشد که در سال‌های ۱۳۹۰ تا ۱۳۹۳ مورد حسابرسی مالیات بر ارزش افزوده قرار گرفته‌اند. نتایج تحقیق نشان می‌دهد دو تکنیک از تکنیک ارزیابی ۷۰ درصد

^۱. Area Under the Curve

برخوردار هستند و نوعی ادغام در نتایج این دو تکنیک توانسته است با کسب نزدیک به ۸۳ درصد صحت ارزیابی از توان بالاتری برخوردار باشد.

۳-۲. پژوهش‌های خارجی

نجوین^۱ (۲۰۲۲) در پژوهشی به تأثیر عوامل غیراقتصادی بر رفتار تمکین مالیات داوطلبانه: مطالعه موردی کوچک و متوسط شرکت‌ها در ویتنام پرداختند. داده‌های مورد استفاده از نتایج نظرسنجی ۳۳۹ حسابدار مالیاتی و مدیران شرکت‌های کوچک و متوسط جمع‌آوری شد. با استفاده از آزمون آلفای کرونباخ، تحلیل عامل اکتشافی و تحلیل رگرسیون خطی به بررسی این موضوع پرداختند. نتایج تحقیق نشان می‌دهد که امکان بازرگانی و حسابرسی مالیاتی، هنجرهای اجتماعی، دانش مالیاتی، هنجرهای شخصی، درک عادلانه بودن سیستم مالیاتی و کیفیت خدمات مالیاتی تأثیر قابل توجهی بر رفتار تمکین داوطلبانه مالیات شرکت‌های کوچک و متوسط در ویتنام دارد. از این میان امکان بازرگانی و حسابرسی مالیاتی قوی‌ترین تأثیر را دارد.

یامن^۲ و همکاران (۲۰۲۱) به بررسی مقایسه‌ای تأثیر کیفیت محیط نهادی در فرار مالیاتی در کشورهای عضو اتحادیه اروپای قدیم (قبل از ۲۰۱۴) و جدید (بعد از ۲۰۰۴) پرداختند. نتایج آنها نشان داد که رابطه بین کیفیت محیط نهادی و فرار مالیاتی بین کشورهای عضو قدیم و جدید متفاوت است. در حالی که کیفیت نظارتی، پاسخگویی، کنترل فساد، ثبات سیاسی و اثربخشی دولت در رسیدن به اهداف تعیین کننده قابل توجهی در میزان فرار مالیاتی در کشورهای قدیم اتحادیه اروپا است، اما اثربخشی دولت‌ها عامل مهم در کشورهای عضو جدید اتحادیه اروپا است.

وانهولودا^۳ و همکاران (۲۰۲۰) در پژوهشی به کشف تقلب مالیات بر ارزش افزوده با تکنیک‌های قابل تشخیص ناهنجاری بدون نظارت (AD) پرداختند و یک روش ارزیابی جدید ارائه دادند که نشانه‌های عملکرد قابل اطمینان را ارائه می‌دهد و تضمین می‌کند که موارد کلاهبرداری به طور مؤثر توسط روشهای پیشنهادی کشف می‌شود.

1. Nguyen

2. Yamen

3. Vanhoeveld

دیدیمو^۱ و همکاران (۲۰۲۰) در پژوهشی تحت عنوان ارائه یک سیستم تعزیزیه و تحلیل دیداری برای پشتیبانی از کشف فرار مالیاتی پرداختند. آنها با استفاده از تکنیک‌های ردهبندی و پیش‌بینی و رگرسیون و با استفاده از الگوریتم SVM، MLP، RF و Log به توصیف سیستم پشتیبانی درخت تصمیم برای کشف فرار مالیاتی بر مبنای زبان دیداری و تکنیک‌های پیشرفته مصورسازی شبکه پرداختند. این سیستم به استفاده کنندگان از آن، این امکان را می‌دهد که گراف‌های زیرمجموعه نمونه‌های مشکوک را با توجه به تطابق با الگوهای موجود رسم کرده و باعث ادغام نتایج و ایجاد شاخص‌هایی برای طبقه‌بندی کردن مالیات دهنده‌گان بر مبنای ریسک مالی شوند. نتایج نشان داد که سیستم پیشنهادی پشتیبانی درخت تصمیم برای کشف فرار مالیاتی مؤثرتر از سایر سیستم‌ها می‌باشد.

لوپز^۲ و همکاران (۲۰۱۹) در پژوهشی کشف تقلب مالیاتی در مورد اظهارنامه مالیات بر درآمد شخصی ثبت شده در اسپانیا، با استفاده از ابزار پیش‌بینی پیشرفته یادگیری ماشین، با استفاده از مدل‌های شبکه عصبی پرسپترون چند لایه پرداختند. نتایج نشان داد که مدل انتخاب شده دارای نرخ کارایی ۸۴/۳ درصد است که حاکی از بهبود در رابطه با سایر مدل‌های مورد استفاده در کشف تقلب مالیاتی است.

راویسانکار^۳ و همکاران (۲۰۱۱) روش‌های مختلف داده کاوی از جمله شبکه عصبی، ماشین‌بردار پشتیبان، برنامه ریزی ژنتیک، شبکه عصبی گروهی مدیریت داده‌ها، رگرسیون لجستیک و شبکه عصبی احتمالی را بر روی ۲۰۲ شرکت چینی به منظور کشف فرار مالیاتی با استفاده و عدم استفاده از انتخاب ویژگی مورد بررسی قرار دادند. در شرایط عدم استفاده از انتخاب ویژگی، شبکه عصبی احتمالی بالاترین دقت را نتیجه داده است. در شرایط استفاده از فرایند انتخاب ویژگی نیز برنامه ریزی ژنتیک و شبکه عصبی احتمالی دقت‌های بالاتر (البته نزدیک به یکدیگر) را نتیجه داده‌اند. روش مورد استفاده برای انتخاب ویژگی روش ساده آماری آزمون تی می‌باشد.

1. Didimo

2. Lopez

3. Ravisankar

بررسی مطالعات که در بالا به مرور آن پرداخته شد، حوزه جامعی از ویژگی‌ها و عوامل فرار مالیاتی را مورد توجه قرار نداده و بیشتر مبتنی بر ویژگی‌های کلان اقتصادی بودند. از سوی دیگر، تاکنون پژوهش‌های اندکی در زمینه فرار مالیاتی با توجه به اطلاعات حسابداری و سایر اطلاعات غیرمالی انجام شده است. در شناسایی این متغیرها در بیشتر مطالعات انجام شده، صرفاً بر اعداد صورت‌های مالی شامل سود خالص، سود ناخالص، فروش کالا، مجموعه دارایی‌ها، بدھی‌ها، حقوق صاحبان سهام و مالیات ابرازی به منظور پیش‌بینی فرار مالیاتی و اجتناب مالیاتی استفاده شده است. با این وجود ویژگی‌های حسابرسان مالیاتی مورد توجه لازم واقع نشده‌اند و حتی به ضرورت اتخاذ رویکرد جامع و کلان با لحاظ تمامی متغیرهای فوق در ارزیابی فرار مالیاتی اشاره نشده است. براین اساس، با توجه به مطالعات بررسی شده این تحقیق از این منظر حائز اهمیت است که با تحلیل مؤلفه‌های مالی و غیرمالی^۱ در سطح اقتصاد کلان، مؤدیان و حسابرسان مالیاتی مدلی جامع جهت پیش‌بینی فرار مالیاتی مؤدیان حقوقی پرداخته است.

۴. طرح مسئله

- با توجه به مبانی نظری و پیشینه تجربی، این پژوهش در صدد یافتن پاسخ به سؤالات زیر می‌باشد:
۱. کدام متغیرها در سطح کلان اقتصادی، مؤدیان مالیاتی و حسابرسان مالیاتی بر فرار مالیاتی مؤدیان حقوقی اداره کل امور مالیاتی مازندران اثر می‌گذارند؟
 ۲. کدام یک از روش‌های الگوریتم درخت تصمیم و رگرسیون چند متغیره خطی، پیش‌بینی بهتری از فرار مالیاتی مؤدیان حقوقی سطح اداره کل امور مالیاتی مازندران را ارائه می‌دهد؟

۱-۴. روش‌شناسی پژوهش

این پژوهش از لحاظ هدف، تحقیقی کاربردی، از لحاظ ماهیت، توصیفی؛ از نظر زمان اجرا، مقطعی؛ از نظر منطق اجرا، قیاسی – استقرایی؛ و از جهت اجرا، از نوع پس رویدادی (از طریق اطلاعات

۱. اطلاعات مالی و غیر مالی داده‌های مربوط به سطح مؤدیان مالیاتی و حسابرسان مالیاتی از صورتهای مالی شرکت، اظهارنامه‌های مالیاتی ارائه شده و سامانه‌های سازمان امور مالیاتی، سایت روزنامه رسمی و ... حاصل شده است.

گذشته) می‌باشد. این تحقیق براساس اطلاعات واقعی موجود در صورت‌های مالی اشخاص حقوقی، اظهارنامه مالیاتی مؤدیان اعم از عملکرد و ارزش افزوده، سایر داده‌های مالی و غیرمالی موجود در پرونده‌های فیزیکی مؤدیان مالیاتی و پرونده‌های پرسنلی حسابرسان بوده و با توجه به اینکه نمونه انتخابی این تحقیق از اشخاص حقوقی بوده و جزو پرونده‌های مالیاتی بزرگ و دارای رتبه ریسک بالا و تأمین کننده عمدۀ وصولی این منبع مالیاتی بوده، با استفاده از روش استقرایی به کل جامعه آماری قابل تعمیم می‌باشد.

۴-۲. جامعه آماری پژوهش

در این پژوهش از اطلاعات کمی و کیفی مؤدیان اشخاص حقوقی دارای پرونده در اداره کل امور مالیاتی استان مازندران که دارای ویژگی‌های زیر باشند، بهره گرفته شده است:

- پرونده‌های مالیاتی مورد بررسی، از نوع مؤدیان مالیاتی اشخاص حقوقی بزرگ سطح اداره کل امور مالیاتی مازندران بوده که دارای رتبه ریسک بالایی^۱ می‌باشد.
- اظهارنامه مالیاتی موضوع ماده ۱۱۰ قانون مالیات‌های مستقیم که حاوی اطلاعات ترازنامه و صورت سود و زیان و سایر اطلاعات مالی و غیرمالی متنکی به دفاتر و اسناد و مدارک خود را در موعد مقرر تسلیم اداره امور مالیاتی محل فعالیت اصلی خود نماید.
- در بازه زمانی ۱۳۹۸-۱۳۹۱ فعالیت داشته باشند. داده‌های پرونده‌های مالیاتی قبل از عملکرد ۱۳۹۱ به دلیل امحاء و بعد از عملکرد ۱۳۹۸ به دلیل عدم قطعی سازی در دسترس نبوده و قابلیت بررسی ندارند.
- رسیدگی پرونده‌ها متنه‌ی به صدور برگ تشخیص و همچنین برگ قطعی مالیات گردیده باشد.

۱. انتخاب رتبه ریسک مؤدیان با معاونت فناوری‌های سازمان امور مالیاتی کشور می‌باشد. معاونت فناوری موظف است بر اساس معیارها و قواعد سنجش ریسک منتخب سازمان که از مدل و مجموعه معیارهای ریسک تأیید و ابلاغ شده توسط معاونت حقوقی و فنی مالیاتی تعیین خواهد شد، سامانه سنجش ریسک مؤدیان را فعل نموده و آنها را جهت قرارگیری مستقیم در انباره حسابرسی ادارات امور مالیاتی ارسال می‌نمایند. قابل ذکر است رتبه بندي ریسک مؤدیان به سه طبقه مؤدیان با ریسک بالا، ریسک متوسط و ریسک پایین تقسیم‌بندی می‌شوند.

- موضوع فعالیت اشخاص حقوقی انتخابی از نوع فعالیت کشاورزی نباشد که منجر به معافیت از پرداخت مالیات براساس ماده ۸۱ قانون مالیاتهای مستقیم گردد.

بعد از در نظر گرفتن محدودیت‌های فوق، تعداد ۹۶۴ پرونده مالیاتی مؤذیان بزرگ حقوقی در دوره زمانی ۱۳۹۸ تا ۱۳۹۱ مورد بررسی قرار گرفت.

در این پژوهش برای جمع‌آوری داده‌ها و اطلاعات از روش‌های کتابخانه‌ای استفاده شده است. مبانی نظری پژوهش از کتب، مجلات و سایر سایت‌های تخصصی فارسی و لاتین و داده‌های مالی و غیر مالی مورد نیاز با مراجعت به ادارات امور مالیاتی استان مازندران و مشاهده داده‌های موجود در پرونده‌های مالیاتی و اخذ سایر اطلاعات مورد نیاز گردآوری گردید.

۴-۳. متغیرهای پژوهش

بر اساس ماده ۲۰۱ قانون مالیات‌های مستقیم هرگاه مؤذی به قصد فرار از مالیات از روی علم و عمد به ترازنامه و حساب سودوزیان یا به دفاتر و اسناد و مدارکی که برای تشخیص مالیات ملاک عمل می‌باشد و برخلاف حقیقت تهیه و تنظیم شده است استناد نماید فرار مالیاتی گفته می‌شود. لذا متغیر وابسته این تحقیق، فرار مالیاتی (Y) است که همانند پژوهش نمازی و صادق‌زاده (۱۳۹۷) و دستگیر و غریبی (۱۳۹۴) و با استناد به ماده فوق الذکر، لحاظ گردیده است. نهایتاً از طریق رابطه ذیل محاسبه خواهد شد.

$$\Delta\% = (TAXIN_{it} - ACCIN_{it}) * 100 \quad (1)$$

که بر این اساس:

$\Delta\%$: درصد اختلاف بین درآمد مشمول مالیات ابرازی (ACCIN) و درآمد مشمول مالیات قطعی شده (TAXIN) شرکت i در سال t

در مورد متغیرهای پیش‌بینی کننده، توضیحی و اثرگذار بر فرار مالیاتی، براساس مطالعات ادبیات و پیشینه تحقیق، مجموعه متنوعی از متغیرها باید در نظر گرفته شوند. از این رو مهمترین

ویژگی‌های متمایز کننده این پژوهش نسبت به مطالعات پیشین، توجه به دامنه گسترده‌ای از انواع متغیرها در کنار یکدیگر می‌باشد. به این ترتیب با مطالعه پژوهش‌های پیشین، پر تکرارترین و مرتبه‌ترین متغیرها در شکل گیری فرار مالیاتی در سطح کلان اقتصادی (کنو (۲۰۲۰)؛ باربوتا (۲۰۰۸)؛ ریچاردسون (۲۰۰۶)؛ کارنه و نورزاد (۱۹۹۰)؛ هیکینز (۱۹۸۵)؛ رضاقلی زاده و همکاران (۱۳۹۸)؛ مطلبی و همکاران (۱۳۹۸)؛ روستا و حیدریه (۱۳۹۳)؛ هادیان و تحولی (۱۳۹۲)؛ صمدی و تابنده (۱۳۹۲)؛ نصرافهانی و همکاران (۱۳۹۱)؛ فلاحتی، و همکاران (۱۳۹۱)؛ امین خاکی و همکاران (۱۳۹۰)؛ سیدزارع (۱۳۸۰)؛ در سطح مؤدیان مالیاتی (روان و همکاران (۲۰۱۹)؛ دارگلاس و همکاران (۲۰۱۹)؛ لاری (۲۰۱۵)؛ یون سوانگ کو و همکاران (۲۰۱۱)؛ زنگانه و همکاران (۱۳۹۹)؛ ثابی پور (۱۳۹۹)؛ جمشیدی و همکاران (۱۳۹۸)؛ خسروی تبار و همکاران (۱۳۹۸)؛ نمازی و صادق زاده مهارلوی (۱۳۹۷)؛ ملکیان و فرزاد (۱۳۹۶)؛ محبوی (۱۳۹۵)؛ ناصری و محمدزاده سالطه (۱۳۹۴)؛ دستگیر و غربی (۱۳۹۴)؛ باقری و عموری (۱۳۸۸) و در سطح حسابرسان مالیاتی (جیوانی و همکاران (۲۰۱۹)؛ گومز آگیوالر و رویز باربادیلو (۲۰۰۶)). وطنپرست و همکاران (۱۴۰۰)؛ اثنی عشری و همکاران (۱۳۹۹)؛ ثابی پور (۱۳۹۹)؛ همتی و همکاران (۱۳۹۸)؛ فلاحتی و علی پور (۱۳۹۷)؛ نقی زهی و شهرزاد محمد خانلی (۱۳۸۹)، جمع‌آوری؛ و با توجه به گزارش‌های مستمر عملکرد سازمان امور مالیاتی، داده‌های موجود در پرونده‌های مؤدیان مالیاتی اعم از مالیات بر درآمد و مالیات بر ارزش افزوده و در کنار نظر کارشناسان خبره امور مالیاتی، متغیرهای پژوهش، نهایی و در قالب جدول (۱) ارائه شده است. لازم به ذکر است که برخی از متغیرها به دلیل همبستگی بالا با سایر متغیرها، حذف شده‌اند. به این ترتیب براساس اطلاعات مندرج در جدول (۱)، در گروه متغیرهای عوامل سطح کلان اقتصادی (Sk) ۱۰ متغیر، گروه متغیرهای سطح مؤدیان مالیاتی (Sm) ۲۷ مورد و گروه متغیرهای سطح حسابرسان مالیاتی (Sh) ۲۰ متغیر مورد توجه قرار گرفته‌اند.

جدول ۳. عوامل مؤثر بر فرار مالیاتی مؤدیان مالیاتی- اشخاص حقوقی

ردیف	گروه اصلی عوامل مؤثر بر فرار مالیاتی	متغیر(عامل مؤثر بر فرار مالیاتی)	شناسه	تعريف عملياتي متغير	منابع استخراج داده ها
۱	عوامل کلان اقتصادادی	ضریب جینی	Sk1	توزع درآمد(یا ثروت) در یک جمعیت	سایت رسمی بانک مرکزی و مرکز آمار ایران
۲		نرخ نورم سالانه	Sk2	ارزش ریال در یک مقطع زمانی نسبت به مقطع زمانی دیگر	
۳		بهای کالا و خدمات مصرفی	Sk3	تغییرات قیمت کالاهای خدمات	
۴		نرخ رشد تولید ناخالص داخلی	Sk4	سرعت رشد یک اقتصاد در طی یکسال	
۵		نرخ بهره	Sk5	نرخ بهره پرداختی با اخذ تسهیلات بانکی	
۶		نرخ رشد ارز	Sk6	نرخ رشد واحد پولی دلار	
۷		قیمت نفت	Sk7	میانگین سالانه قیمت هر بشکه نفت اوپیک به دلار	
۸		قیمت طلا	Sk8	میانگین سالانه قیمت طلای ۱۸ عیار	
۹		نرخ سود بانکی	Sk9	نرخ سود دریافتی سپرده‌های بانکی	
۱۰		تراز تجاری	Sk10	جریان صادرات و واردات در یک سال	

ردیف	گروه اصلی عوامل مؤثر بر فرار مالیاتی	متغیر(عامل مؤثر بر فرار مالیاتی)	شناسه	تعریف عملیاتی متغیر	منابع استخراج داده ها
۱۱	ویژگی های مؤیدان مالیاتی	شرکت های خانوادگی	Sm1	در صورت وابستگی اعضای هیات مدیره عدد یک و در غیر این صورت عدد صفر لحاظ می گردد.	
۱۲		زیان دهی شرکت (نمایش زیانهای متوالی سالانه در اظهارنامه ها با وجود تداوم فعالیت عادی)	Sm2	در صورت وجود زیان در صورتهای مالی دو سال متوالی گذشته عدد یک و در غیر این صورت عدد صفر لحاظ می گردد.	
۱۳		نسبت بهای مواد خارجی	Sm3	نسبت مبلغ کالای خارجی خریداری شده به کل کالای خریداری شده اعم از داخلی و خارجی	
۱۴		معامله خرید با اشخاص وابسته	Sm4	نسبت مبلغ خرید از اشخاص وابسته به کل خرید طی دوره	
۱۵		معامله فروش با اشخاص وابسته	Sm5	نسبت مبلغ فروش به اشخاص وابسته به کل فروش طی دوره	
۱۶		نسبت جاری شرکا	Sm6	نسبت جاری شرکا به کل داراییها	
۱۷		نسبت هزینه حقوق	Sm7	نسبت هزینه حقوق کارکنان به مبلغ فروش	
۱۸		نسبت مواد مصرفی	Sm9	نسبت مواد مصرفی به فروش کالا و خدمات	

منابع استخراج داده ها	تعریف عملیاتی متغیر	شناسه	متغیر(عامل مؤثر بر فرار مالیاتی)	گروه اصلی عوامل مؤثر بر فرار مالیاتی	ردیف
دادهای موجود در اظهارنامه های مالیاتی (اعم از عملکرد و ارزش افزوده)، صورتهای مالی شرکت ها و پسایر اطلاعات موجود پرونده فیزیکی مؤذین مالیاتی	نسبت اطلاعات فروش موجود در سامانه ۱۶۹ ق.م به کل مبلغ فروش طی دوره	Sm10	نسبت اطلاعات مالی مربوط به فروش در سامانه ۱۶۹ ق.م.	۱۹	
	نسبت اطلاعات خرید موجود در سامانه ۱۶۹ ق.م به مبلغ خرید طی دوره	Sm11	نسبت اطلاعات مالی مربوط به خرید در سامانه ۱۶۹ ق.م.	۲۰	
	اگر مبلغ مالیات تشخیص شده برای اعضای هیات مدیره یا کارکنان در راستای رسیدگی به پرونده تراکنش های مشکوک بانکی رسیدگی کرده باشد عدد یک و در غیر صفر لحاظ می گردد.	Sm12	وجود تراکنش بانکی مشکوک	۲۱	
	نسبت اعتبار مالیاتی ارزش افزوده (خرید و کالا و خدمات مشمول مالیات) تأیید نشده به اعتبار ابرازی مؤذی	Sm13	نسبت اعتبار مالیاتی	۲۲	
	نسبت معامله انجام شده (فروش) به نام کارکنان یا اعضای هیات به فروش کل دوره جاری	Sm14	تنظيم معاملات به نام دیگران		۲۳

ردیف	گروه اصلی عوامل مؤثر بر فرار مالیاتی	متغیر(عامل مؤثر بر فرار مالیاتی)	شناسه	تعریف عملیاتی متغیر	منابع استخراج داده ها
۲۴	سود سپرده		Sm15	نسبت مبلغ سود سپرده بانکی به مبلغ فروش کالا طی دوره	
۲۵	خودمصرفی کالا		Sm16	نسبت اعتبار مالیاتی ارزش افزوده (بابت خرید کالا و خدمات مشمول مالیات) تأیید شده با بت عرضه به خود به کل اعتبار ابرازی	
۲۶	تبانی در معاملات		Sm17	در صورت وجود تبانی در معاملات(خرید و فروش به قیمت کمتر یا بیشتر از ارزش منصفانه) عدد یک و در غیر این صورت عدد صفر لحاظ می گردد.	
۲۷	خریدهای صوری		Sm18	نسبت مبلغ خریدهای صوری از شرکت های فقد اعتبار مالیاتی به کل خرید طی دوره	
۲۸	نسبت سود ناخالص به فروش		Sm19	نسبت مبلغ سود ناخالص به مبلغ فروش طی دوره	
۲۹	نسبت سود خالص به فروش		Sm20	نسبت مبلغ سود خالص به مبلغ فروش کالا طی دوره	
۳۰	نرخ بازده دارایی ها		Sm21	نسبت مبلغ سود خالص به مبلغ جمع دارایی ها	
۳۱	نسبت هزینه تبلیغات به فروش		Sm22	نسبت مبلغ هزینه تبلیغات به مبلغ فروش طی دوره	
۳۲	نسبت اهرم مالی		Sm23	نسبت مجموع بدھی ها به دارایی ها	

منابع استخراج داده ها	تعریف عملیاتی متغیر	شناسه	متغیر(عامل مؤثر بر فرار مالیاتی)	گروه اصلی عوامل	ردیف مؤثر بر فرار مالیاتی
	در صورت انجام تکاليف قانونی در راستای ماده ۹۵ قانون مالیاتهای مستقیم عدد یک و در غیر این صورت عدد صفر لحاظ می گردد.	Sm24	انجام تکلیف قانونی مطابق ماده ۹۵ قانون مالیاتهای مستقیم		۳۳
	در صورتی که براساس سطح اهمیت پرونده مالیاتی رسیدگی آن توسط حداقل دو حسابرس ارشد مالیاتی انجام شود عدد یک و غیر عدد صفر لحاظ می شود.	Sm25	نحوه رسیدگی به پرونده های مالیاتی		۳۴
	نسبت استفاده از کارت بازرگانی دیگران برای انجام خریدهای خارجی به کل مبلغ خرید مواد خارجی	Sm26	استفاده از کارت بازرگانی دیگران		۳۵
	در صورتی که پرونده مذکور در سال مورد رسیدگی جزو صد پرونده بزرگ سطح اداره کل باشد عدد یک و غیر عدد صفر لحاظ می شود.	Sm27	اندازه مؤذی		۳۶
	در صورتی که صدور برگ تشخیص مالیات عملکرد پرونده مذکور بر اساس سامانه های نرم افزاری جدید (اریس) انجام گیرد عدد یک و غیر عدد صفر لحاظ می گردد.	Sm28	سامانه صدور برگ تشخیص		۳۷

ردیف	گروه اصلی عوامل مؤثر بر فرار مالیاتی	متغیر(عامل مؤثر بر فرار مالیاتی)	شناسه	تعریف عملیاتی متغیر	منابع استخراج داده ها
۳۸	ویژگی های حسابرسان (ممیزان) مالیاتی	سن سرممیز	Sh1	سن رئیس گروه حسابرسی مالیاتی در سال مورد رسیدگی	پرونده پرسنلی حسابرسان مالیاتی، احکام رسیدگی پرونده ها و سایر اطلاعات موجود در پرونده های فیزیکی مؤدیان
۳۹		متوسط سن گروه رسیدگی پرونده در سال مورد رسیدگی (شامل رئیس گروه حسابرسی مالیاتی و حسابرسان ارشد مالیاتی)	Sh2		
۴۰		تجربه سرممیز	Sh3	سابقه کاری رئیس گروه حسابرسی مالیاتی در سال مورد رسیدگی	
۴۱		تجربه اعضای گروه رسیدگی	Sh4	متوسط سابقه کاری اعضای گروه (از بدو استخدام) در سال مورد رسیدگی (شامل رئیس گروه حسابرسی مالیاتی و حسابرسان ارشد مالیاتی)	
۴۲		وجود حسابرس خانم در گروه رسیدگی	Sh5	در صورت وجود حسابرس با جنسیت خانم در گروه رسیدگی عدد یک و در غیر این صورت عدد صفر لحاظ می گردد.	
۴۳		تخصص مالی سرممیز	Sh6	در صورتی که مدرک و رشته تحصیلی رئیس گروه حسابرسی مالیاتی، حسابداری باشد عدد یک و در غیر این صورت عدد صفر لحاظ می گردد.	
۴۴		تخصص مالی اعضای گروه رسیدگی	Sh7	نسبت اعضای گروه با مدرک حسابداری به کل اعضا (شامل رئیس گروه حسابرسی مالیاتی و حسابرسان ارشد مالیاتی)	

منابع استخراج داده ها	تعریف عملیاتی متغیر	شناسه	متغیر(عامل مؤثر بر فرار مالیاتی)	گروه اصلی عوامل مؤثر بر فرار مالیاتی	ردیف
	چنانچه سرممیز بیش از یکسال رسیدگی پرونده مالیاتی شرکت را انجام داده باشد، عدد یک و در غیر این صورت عدد صفر لحاظ می گردد.	Sh8	حسابرس نخستین		۴۵
	نسبت تفاضل مالیات ابرازی از مالیات تشخیصی به تفاضل مالیات قطعی از مالیات تشخیصی	Sh9	کیفیت رسیدگی		۴۶
	در صورتی که رئیس گروه حسابرسی در سال های قبل در صنعت مشابه رسیدگی داشته عدد یک و در غیر این صورت عدد صفر لحاظ می گردد.	Sh10	تخصص در صنعت سرممیز		۴۷
	نسبت سابقه کاری در پست سازمانی رئیس گروه حسابرسی مالیاتی (از تاریخ انتصاب به سمت مذکور) به کل سابقه کاری وی از بدو استخدام	Sh11	رضایت شغلی سرممیز		۴۸
	نسبت سابقه کاری در پست سازمانی حسابرس ارشد مالیاتی اول (از تاریخ انتصاب به سمت مذکور) به کل سابقه کاری وی از بدو استخدام	Sh12	رضایت شغلی حسابرس ارشد اول		۴۹

منابع استخراج داده ها	تعریف عملیاتی متغیر	شناسه	متغیر(عامل مؤثر بر فرار مالیاتی)	گروه اصلی عوامل مؤثر بر فرار مالیاتی	ردیف
	نسبت سابقه کاری در پست سازمانی حسابرس ارشد مالیاتی دوم (از تاریخ انتصاب به سمت مذکور) به کل سابقه کاری وی از بدو استخدام	Sh13	رضایت شغلی حسابرس ارشد دوم		۵۰
	نسبت بررسی اطلاعات واصله از طرف دوم معاملات (خرید دیگران از شرکت مورد رسیدگی) به کل اطلاعات واصله از سامانه ۱۶۹ ق.م. میزان بررسی یارهگیری اطلاعات با دفاتر مودی	Sh14	استفاده حسابرسان از اطلاعات فروش از سامانه ۱۶۹ ق.م.		۵۱
	نسبت بررسی اطلاعات واصله از طرف دوم معاملات (فروش دیگران از شرکت مورد رسیدگی) به کل اطلاعات واصله از سامانه ۱۶۹ ق.م.	Sh15	استفاده حسابرسان از اطلاعات خرید از سامانه ۱۶۹ ق.م.		۵۲
	نسبت حسابها و استناد دریافتی بررسی شده به کل حسابها و استناد دریافتی	Sh16	راهبرد حسابرسان مالیاتی (حسابها و استناد دریافتی)		۵۳
	نسبت حسابها و استناد پرداختی بررسی شده به کل حسابها و استناد پرداختی	Sh17	راهبرد حسابرسان مالیاتی (حسابها و استناد پرداختی)		۵۴

ردیف	گروه اصلی عوامل مؤثر بر فرار مالیاتی	متغیر(عامل مؤثر بر فرار مالیاتی)	شناسه	تعریف عملیاتی متغیر	منابع استخراج داده ها
۵۵	دوره تصدی سرمیز		Sh18	تعداد سال‌های رسیدگی رئیس گروه حسابرسی مالیاتی به پرونده مذکور از سال ۱۳۸۹ تا سال مورد بررسی توسط حسابرس	
۵۶	فشارکاری حسابرس		Sh19	$WS = \frac{\sum_{i=1}^m \sum_{j=1}^n TA(ij)}{m}$ WS: میانگین فشارکاری حسابرس در پرونده TA: لگاریتم طبیعی جمع دارانی های شرکت $\hat{}$ حسابرسی شده به وسیله حسابرس، n : کل شرکت های حسابرسی شده به وسیله حسابرس ا در سال مالی و m : تعداد امضاهای حسابسان بر روی گزارش حسابرسی مالیاتی شرکت $\hat{}$.	
۵۷	چند هدفی بودن حسابرس رسیدگی همزمان چند منبع مالیاتی مؤذدی(عملکرد، ارزش افزوده، حقوق، تکایفی و ۱۶۹ ق.م) توسط حسابسان در یکسال		Sh20	نسبت تعداد نقش‌های اجرا شده به کل نقش‌های موجود توسط رئیس گروه حسابرسی مالیاتی	

مأخذ: یافته‌های محقق

۵. تجزیه و تحلیل داده ها

به منظور پیش‌بینی فرار مالیاتی، ابتدا با استفاده از الگوریتم‌های انتخاب ویژگی، تأثیر گذارترین ویژگی‌ها از بین دامنه گسترده‌ای از شاخص‌های سطح کلان اقتصادی، مؤدیان و حسابرسان مالیاتی انتخاب خواهد شد. سپس برای طراحی مدل از الگوریتم درخت تصمیم و رگرسیون استفاده شده است. در این تحقیق به منظور استخراج ویژگی از الگوریتم بهینه‌سازی سینوس کسینوس بهره برده شده است.

۱-۵. الگوریتم بهینه سازی سینوس کسینوس

الگوریتم بهینه‌سازی سینوس کسینوس، یک الگوریتم فرامکاشفه‌ای^۱ مبتنی بر جمعیت است که برای حل مسائل بهینه‌سازی پیوسته معرفی شده است. این الگوریتم با توزیع تصادفی راه حل‌ها در فضای جستجو شروع می‌شود. پس از محاسبه تابع برازش^۲ هر راه حل، راه حل با بهترین برازش به عنوان راه حل نهایی در نظر گرفته می‌شود (حافظ^۳ و همکاران، ۲۰۱۶). راه حل نهایی در معادله بهروزرسانی موقعیت که با معادله زیر نشان داده شده است استفاده می‌شود و توسط آن موقعیت سایر راه حل‌ها تغییر می‌کند.

$$X_i^{t+1} = \begin{cases} X_i^t + r_1 \times \sin(r_2) \times |r_3 P_i^t - X_i^t|, & r_4 < 0.5 \\ X_i^t + r_1 \times \cos(r_2) \times |r_3 P_i^t - X_i^t|, & r_4 \geq 0.5 \end{cases} \quad (2)$$

که در آن X_i مقدار بعد i -امین جواب فعلی در تکرار t است، P_i^t موقعیت راه حل مقصد در بعد i -ام و تکرار t -ام است و r_1 و r_2 و r_3 اعداد تصادفی هستند. در این معادله، یک فرمول با تولید عدد تصادفی r_4 با توزیع یکتواخت بین ۰ و ۱ انتخاب می‌شود. الگوریتم بهینه‌سازی سینوس کسینوس تا زمانی که معیارهای خاتمه برآورده شود اجرا می‌شود.

1. Meta heuristic
2. Fitness function
3. Hafez

الگوریتم سینوس اکتشاف^۱ و بهره‌برداری^۲ را با استفاده از سه پارامتر r_1 ، r_2 و r_3 کنترل و راه حل‌ها را به موقعیت بعدی هدایت می‌کند. پارامتر r_1 توانایی ایجاد تعادل بین اکتشاف و بهره‌برداری را در مراحل ابتدا و انتهای الگوریتم دارد. این پارامتر جهت راه حل جدید را به سمت خارج از راه حل مقصد تعیین می‌کند. این پارامتر فرآیند جستجو را برای کاوش کل فضای جستجو حتی دور از راه حل مقصد در مراحل اولیه الگوریتم یا برای بهره‌برداری در نزدیکی راه حل مقصد به منظور یافتن راه حل‌های بهتر در آخرین مراحل الگوریتم هدایت می‌کند. اگر $0 < r_1 < r_2$ فاصله بین راه حل و راه حل مقصد کاهش می‌یابد، در حالی که اگر $0 < r_1 < r_2$ باشد، این فاصله افزایش پیدا می‌کند. پارامتر r_2 با در نظر گرفتن حداکثر تکرار T ، تکرار فعلی a و مقدار ثابت a محاسبه می‌شود که در معادله زیر نشان داده شده است.

$$r_1 = a - t \frac{a}{T_{max}} \quad (3)$$

پارامتر تصادفی r_3 مقدار فاصله راه حل فعلی را از موقعیت راه حل مقصد نشان می‌دهد. مقدار بالاتر این پارامتر منجر به اکتشاف می‌شود زیرا فاصله بین راه حل فعلی و راه حل مقصد بیشتر است، در حالی که مقدار کمتر نشان دهنده فاصله کمتر است و منجر به بهره‌برداری می‌شود. سومین پارامتر r_3 پارامتر وزنی برای نشان دادن تأثیر راه حل مقصد در تعریف فاصله است. شبه کد این الگوریتم به صورت زیر است:

-
1. Exploration
 2. Exploitation

```

begin SCA Algorithm Main

    Randomly initialize each search agents in the
    population Xi(i = 1, 2, ..., m);

    Initialize the max numbers of iteration Tmax;

    while T < Tmax do for each iteration

        foreach search agent Xi in the population do

            Evaluate Xi via the fitness function. if f(Xi)
            better than f(P) then

                Set P = Xi;

            end

        end

        Update r1 using equation 2.

        Generate randomly new values for r2, r3 and s.

        foreach search agent Xi in the population do

            Update Xi using equation 1.

        end

        end

    end

    Return P best solution obtained so far.
end

```

به منظور پیاده سازی روش مورد نظر از نرم افزار متلب استفاده شده است و تابع این روش با نام `SCA.m` باشد.^۱ در ادامه ورودی و خروجی های این الگوریتم بیان شده است:

1. <https://www.mathworks.com/matlabcentral/fileexchange/80671-sine-cosine-algorithm-for-feature-selection>

```
%-----%
%           % Sine Cosine Algorithm (SCA) source code
%-----%
%---Inputs-
%   % feat      : feature vector ( Instances x Features )
%   % label     : label vector ( Instances x 1 )
%   % N         : Number of solutions
%   % max_Iter  : Maximum number of iterations
%   % alpha     : Constant
%---Output-
%   % sFeat     : Selected features (instances x features)
%   % Sf        : Selected feature index
%   % NF        : Number of selected features
%   % curve     : Convergence curve
%-----%
```

در این پژوهش بعد از اجرای الگوریتم بهینه‌سازی سینوس کسینوس از بین ۵۷ شاخص‌های فرار مالیاتی در سطح کلان اقتصادی، مؤیدان مالیاتی و حسابرسان مالیاتی ۸ متغیر به عنوان ویژگی‌های کارآمدتر به شرح جدول ۴ به دست آمده است:

جدول ۴. متغیرهای استخراجی با استفاده از الگوریتم سینوس کسینوس

متغیرها	شناسه
نسبت مواد مصرفی	SM9
نسبت اعتبار مالیاتی	SM13
نسبت سود ناخالص	SM19
نسبت سود خالص	SM20
انجام تکالیف قانونی	SM24
روش رسیدگی به پرونده مالیاتی	SM25
استفاده از اطلاعات فروش در سامانه معاملات فصلی (۱۶۹ ق.م.)	SH14
استفاده از اطلاعات خرید در سامانه معاملات فصلی (۱۶۹ ق.م.)	SH15

مأخذ: یافته‌های محقق

در این بخش نتایج آزمون‌های آماری از طریق آمار توصیفی و سپس آمار استنباطی و نتایج رگرسیون و الگوریتم درخت تصمیم بیان گردیده است.

آمار توصیفی

در جدول شماره ۲ آمار توصیفی متغیرهای استخراجی با استفاده از الگوریتم سینوس-کسینوس ارائه شده است. در این تحقیق ۸ متغیر در دو سطح مؤدیان مالیاتی (Sm) و حسابرسان مالیاتی (Sh) به عنوان متغیرهای پیش‌بینی کننده فرار مالیاتی (Y) انتخاب شده‌اند. به منظور تشخیص نرمال بودن داده‌های مورد استفاده از آماره جارک بر^۱ استفاده می‌شود. همان‌طور که در جدول ۵ مشاهده می‌شود، برای تمامی متغیرهای مستقل و وابسته آماره جارک برآ نزدیک به صفر است و این نشان می‌دهد که این متغیرها از توزیع نرمال پیروی می‌کنند.

جدول ۵. آمار توصیفی متغیرهای پژوهش

نام متغیر	حداقل	حداکثر	میانگین	انحراف معیار	آماره جارک برآ
۷	۰.....	۰.۲۲۲	۰.۰۱۳	۰۰۲۰	۰۰۰۱
sm9	۰..۰۱۷	۰.۸۱۷	۰.۰۵۱۲	۰۰۰۱۸۵	۰۰۰۱
sm13	۰.....	۱.۰۰۰	۰.۰۲۰۵	۰۰۰۰۲۵۸	۰۰۰۱
sm19	-۰.۴۵۹	۰.۵۶۲	۰.۰۰۷۵	۰۰۰۰۱۳۰	۰۰۰۱
sm24	۰.....	۱.۰۰۰	۰.۰۷۹۱	۰۰۰۰۴۰۶	۰۰۰۱
sm25	۰.....	۱.۰۰۰	۰.۰۸۹۹	۰۰۰۰۳۰۱	۰۰۰۱
sh14	۰.....	۱.۰۰۰	۰.۰۷۵۵	۰۰۰۰۳۴۸	۰۰۰۱
sh15	۰.....	۱.۰۰۰	۰.۰۸۴۱	۰۰۰۰۲۸۲	۰۰۰۱

مأخذ: یافته‌های پژوهش

در این تحقیق به منظور مقایسه تطبیقی روش‌های سنتی و نوین پیش‌بینی کننده فرار مالیاتی به ترتیب از دو روش الگوریتم درخت تصمیم و رگرسیون خطی ساده استفاده شده است.

۱. این آزمون توسط کارلوس جارک اقتصاددان مکزیکی و آنیل برا اقتصاددان آمریکایی در سال ۱۳۷۸ در مقاله‌ای با عنوان آزمون نرمال بودن مشاهدات و باقی مانده‌های رگرسیون معرفی شد. به کمک این آزمون آماره آن می‌توانیم نرمال بودن داده‌ها را مورد بررسی قرار دهیم.

۵-۲. الگوریتم درخت تصمیم

درخت‌های پیش‌بینی از یک درخت برای نمایش تقسیم‌بندی‌های بازگشتی استفاده می‌کنند. هر یک از گره‌های پایانی یا برگ‌های درخت نمایانگر یک سلول از تقسیم‌بندی است که یک مدل به آن متصل شده و فقط در آن سلول اعمال می‌شود. اگر x در سلول مربوطه تقسیم‌بندی باشد، نقطه x متعلق به یک برگ است. برای اینکه متوجه شویم در کدام سلول هستیم، از گره ریشه درخت شروع کرده و دنباله‌ای از سوالات را در مورد ویژگی‌ها می‌پرسیم. گره‌های داخلی درخت با سوالات برچسب گذاری می‌شوند و لبه‌ها یا شاخه‌های بین آنها با پاسخ برچسب گذاری می‌شوند. اینکه کدام سوال بعدی را پرسیم بستگی به پاسخ سوالات قبلی دارد (سنها و گنجیل، ۲۰۱۹).

مزیت‌های درخت تصمیم به صورت زیر است:

- پیش‌بینی سریع است (بدون محاسبات پیچیده، فقط به دنبال ثابت‌ها در درخت).
- در ک اینکه چه متغیرهایی در پیش‌بینی مهم هستند آسان است (به درخت نگاه کنید).
- اگر برخی از داده‌ها از دست رفته باشد، ممکن است نتوانیم تمام مسیر درخت را تا یک برگ طی کنیم، اما همچنان می‌توانیم با میانگین‌گیری تمام برگ‌های درخت فرعی که به آن می‌رسیم، پیش‌بینی کنیم.
- مدل یک پاسخ ناهموار^۱ می‌دهد، بنابراین زمانی که سطح رگرسیون واقعی صاف نیست، می‌تواند کار کند. با این حال، اگر صاف باشد، سطح تکه‌ای ثابت می‌تواند به طور دلخواه به آن نزدیک شود (با برگ‌های کافی).
- الگوریتم‌های سریع و قابل اعتمادی برای یادگیری این درختان وجود دارد.

الگوریتم پایه درخت رگرسیون به شرح زیر است:

1. با یک گره منفرد که شامل همه داده‌ها است شروع کنید. واریانس آن گره را محاسبه کنید.

1. jagged response

۲. اگر تمام نقاط گره برای همه متغیرهای مستقل مقدار یکسانی دارند، توقف کنید. در غیر این صورت، در تمام تقسیمات باینری همه متغیرها جستجو کنید تا واریانس را تا حد امکان کاهش دهد. اگر بزرگترین کاهش در واریانس کمتر از یک حد آستانه باشد یا یکی از گرههای حاصل دارای تعداد داده کمی باشد، توقف کنید. در غیر این صورت، آن تقسیم‌بندی را انجام دهید و دو گره جدید ایجاد کنید.

۳. در هر گره جدید، به مرحله ۱ برگردید.

در درخت تصمیم با تقسیم متوالی فضای ویژگی به صورت مجموعه‌ای از مستطیل‌ها و سپس قرار دادن یک مدل ساده در هر یک از این مستطیل‌ها و با استفاده از داده‌های ورودی، مدل‌سازی انجام می‌شود. روش درخت تصمیم با در نظر گرفتن داده‌های ورودی و تمام تقسیمات دو دوبی ممکن بر روی هر پیش‌بینی کننده شروع می‌شود، و سپس انتخاب تقسیم براساس بهترین معیار بهینه‌سازی انجام می‌شود. این عمل به صورت بازگشتی تکرار می‌شود تا زمانی که شرط توقف برآورده شود. مدل‌های رگرسیونی یا طبقه‌بندی توسط درخت تصمیم به شکل ساختار درختی ساخته می‌شوند. در این ساختار درختی، مجموعه نمونه‌ها، به زیر مجموعه‌های کوچک‌تر تجزیه می‌شوند و یک درخت تصمیم مرتبط با نمونه‌ها، به صورت تدریجی توسعه می‌یابد. نتیجه نهایی، یک درخت با گره‌های تصمیم‌گیری و گره‌های برگ است. یک گره تصمیم، دارای دو یا چند شاخه است که هر یک ارزشی را برای ویژگی مورد آزمایش نشان می‌دهند و گره برگ، تصمیمی را در مورد هدف نهایی نشان می‌دهد.

خطای میانگین مربعات وزنی (WMSE) برای پاسخ‌های موجود در گره t با استفاده از رابطه^۳ محاسبه می‌شود:

$$\varepsilon_t = \sum_{j \in T} w_j (y_j - \bar{y}_t)^2 \quad (4)$$

که در بالا w_j وزن مشاهدات ز و T مجموعه کلیه شاخص‌های مشاهده شده، در گره t می‌باشد. احتمال اینکه یک مشاهده در گره t باشد با استفاده از رابطه^۴ تخمین زده می‌شود.

$$(T) = \sum_{j \in T} w_j P \quad (5)$$

x_i ‌ها به صورت صعودی مرتب می‌شوند. هر عنصر پیش‌بینی مرتب شده یک نامزد تقسیم یا نقطه برش است. تمام شاخص‌های مربوط به مقادیر نامشخص در مجموعه TU ثبت می‌شود. بهترین روش تقسیم گره t با استفاده از x_i با حداکثر رساندن مقدار MSE از بین تمام نامزدهای تقسیم کننده، تعیین می‌شود. یعنی برای همه گره‌های نامزدها تقسیم در x_i داریم:

مشاهدات در گره t به گره‌های چپ و راست تقسیم می‌شود. (به ترتیب t_L و t_R) مقدار MSE محاسبه می‌شود. فرض کنید که برای یک نامزد تقسیم کننده خاص، t_L و t_R به ترتیب دارای شاخص‌های مشاهده در مجموعه‌های TL و TR هستند. اگر x_i حاوی مقادیر از دست رفته نباشد، کاهش MSE برای نامزد تقسیم فعلی به صورت رابطه ۵ است:

$$\Delta I = P(T)\varepsilon_t - P(T_L)\varepsilon_{t_L} - P(T_R)\varepsilon_{t_R} \quad (6)$$

اگر x_i ‌حاوی مقادیر از دست رفته باشد، با فرض اینکه مشاهدات به صورت تصادفی از دست می‌روند، کاهش در MSE به صورت رابطه ۶ است:

$$\Delta I_U = P(T - TU)\varepsilon_t - P(T_L)\varepsilon_{t_L} - P(T_R)\varepsilon_{t_R} \quad (7)$$

که TU - مجموعه‌ای از تمام شاخص‌های مشاهده در گره t است که از دست نرفته است. در نهایت نامزدی انتخاب می‌شود که بیشترین کاهش MSE را داشته باشد.

۳-۵. روش رگرسیون خطی چند متغیره

به عنوان تکنیک دوم از مدل رگرسیون خطی ساده استفاده خواهد شد. در این پژوهش اثرات سال و صنعت در برآذش مدل رگرسیونی کنترل شده است، لذا مدل به صورت مقطعی برآذش خواهد شد و نیازی به آزمون‌های تشخیصی (همانند لیمر و هاسمن) جهت تفکیک اثرات مشترک، ثابت یا تصادفی نخواهد بود. ابتدا پیش‌فرض‌های رگرسیون کلاسیک همچون همخطی بین متغیرهای توضیحی، ناهمسانی واریانس بین جملات خطای خودهمبستگی مرتبه اول بین جملات خطای مورد بررسی قرار خواهند گرفت. در نهایت تخمین نهایی مدل به روش رگرسیون مقاوم (مستحکم) انجام خواهد شد تا مشکلاتی همچون ناهمسانی واریانس و وجود همبستگی مرتبه اول در سطح

شرکت‌های مورد بررسی مرتفع گردد. نتیجه نهایی برآورد مدل رگرسیونی فرار مالیاتی براساس متغیرهای مؤثر انتخاب شده توسط الگوریتم سینوس کسینوس در جدول ۶ ارائه می‌گردد.

جدول ۶. نتایج نهایی برآورد مدل رگرسیونی فرار مالیاتی

شناسه	متغیر	ضریب	خطای استاندارد	سطح معناداری	عامل تورم واریانس
SM9	نسبت مواد مصرفی	۰/۰۱۹۴	۰/۰۰۵۳	۰/۰۰۰۰۰	۱/۲۶
SM13	نسبت اعتبار مالیاتی	۰/۰۱۰۹	۰/۰۰۴۰	۰/۰۰۰۸	۱/۱۳
SM19	نسبت سود ناخالص	۰/۰۰۹۷	۰/۰۰۸۱	۰/۰۰۲۳۲	۱/۷۱
SM20	نسبت سود خالص	۰/۰۰۰۵۰	۰/۰۰۰۶۰۳	۰/۰۰۴۰۱	۱/۶۱
SM24	انجام تکالیف قانونی	۰/۰۰۰۹۷	۰/۰۰۰۲۲	۰/۰۰۰۰۰	۱/۱۱
SM25	روش رسیدگی به پرونده مالیاتی	۰/۰۰۰۹۶	۰/۰۰۰۳۷	۰/۰۰۱۰	۱/۱۷
SH14	استفاده از اطلاعات فروش در سامانه معاملات فصلی (۱۶۹ ق.م)	۰/۰۰۰۷۶	۰/۰۰۰۳۱	۰/۰۱۶	۱/۴۴
SH15	استفاده از اطلاعات خرید در سامانه معاملات فصلی (۱۶۹ ق.م)	۰/۰۰۰۴۸	۰/۰۰۰۴۱	۰/۰۰۲۳۹	۱/۴۲
	عرض از مبدا	۰/۰۳۸۸	۰/۰۰۰۵۷	۰/۰۰۰۰۰	۱/۵۳
	اثرات سال		کنترل شد		
	اثرات صنعت		کنترل شد		
	ضریب تعیین		۰/۱۸۵۴		
	F آماره		۷/۳۰۰۸۲		
	F سطح معناداری		۰/۰۰۰		
	آماره (معناداری) بروش پاگان		(۰۰۰۰۰) (۶۴۳/۶۴)		
	آماره (معناداری) وولریچ		(۰۰۰۴۴) (۴/۱۴۱)		
	خطای مدل (RMSE)		۰/۱۸۵۴		

مأخذ: یافته‌های پژوهش

براساس جدول فوق، متغیرنسبت اعتبار مالیاتی دارای تأثیر مثبت و معنadar و متغیرهای نسبت مواد مصرفی، انجام تکلیف قانونی، روش رسیدگی به پرونده مالیاتی و استفاده از اطلاعات فروش در سامانه معاملات فصلی (۱۶۹ ق.م) تأثیر منفی و معنadarی بر فرار مالیاتی دارند. هم‌چنین مقدار و سطح معنی‌داری آماره F مدل به ترتیب ۷/۳۰۰۸۲ و ۰/۰۰۰۰۰ می‌باشد و بیانگر آن است که مدل تحقیق

به خوبی برآش شده و از معنی‌داری مناسبی برخوردار می‌باشد. مقدار ضریب تعیین مدل برابر با 0.1854 می‌باشد و بیانگر آن است که $18/54$ درصد از تغییرات متغیر وابسته توسط متغیرهای مستقل تحقیق تبیین می‌شود.

جهت سنجش نتایج به دست آمده و مقایسه عملکرد هر یک از روش‌ها برای پیش‌بینی مقادیر فرار مالیاتی با خطای کمتر، یافته‌های حاصل شده از ارزیابی‌ها را با یکدیگر مورد مقایسه قرار دادیم. به منظور محاسبه خطای مربع میانگین مربع خطای استفاده شده است. برای محاسبه میانگین مربع خطاهای، تفاوت بین مقادیر پیش‌بینی شده و مقادیر مشاهده شده مربوطه هر کدام مجنوز می‌شوند و سپس میانگین این خطاهای محاسبه می‌شود و در نهایت جذر میانگین گرفته می‌شود. در این مقایسه مقادیر RMSE هر یک از روش‌ها جهت تحلیل مورد استفاده قرار می‌گیرند.

جدول ۷. مقادیر میانگین مربع خطاهای در دو تکنیک مورد استفاده

روش مورد استفاده	RMSE
الگوریتم درخت تصمیم	۰.۰۱۴۰
رگرسیون چندمتغیره خطی	۰.۰۱۸۵۴

مأخذ: یافته‌های پژوهش

۶. بحث و نتیجه گیری

در این پژوهش یک الگوی پیش‌بینی فرار مالیاتی اشخاص حقوقی ارائه گردید، که بر اساس جستجوی انجام شده و خلاصهای پژوهش‌های داخلی و خارجی تاکنون پژوهشی در این زمینه انجام نشده بود و این کار برای اولین بار اداره کل امور مالیاتی استان مازندران اجرا گردید. در این پژوهش عوامل مؤثر بر فرار مالیاتی در سه سطح کلان اقتصادی، مؤدیان مالیاتی و حسابرسان مالیاتی مورد بررسی و آزمون قرار گرفت؛ در این الگوی پیش‌بینی فرار مالیاتی با دسترسی به ۹۶۴ پرونده مالیاتی اشخاص حقوقی در سطح اداره کل امور مازندران برای عملکرد ۱۳۹۱ لغایت ۱۳۹۸ دنبال شد. بر

1. Root Mean Square Error

این اساس ابتدا تعداد ۵۷ شاخص مؤثر بر فرار مالیاتی اعم از متغیرهای مالی و غیرمالی که تعریف عملیاتی و قابلیت اندازه‌گیری برای آنها وجود داشته مشخص و به منظور آزمون فرضیه استخراج با استفاده از الگوریتم انتخاب ویژگی سینوس کسینوس ۸ متغیر به عنوان ویژگی کارآمدتر انتخاب شدند. از طریق الگوریتم درخت تصمیم و رگرسیون چند متغیره خطی با یکدیگر مقایسه گردیدند.

نتایج حاصل از انجام این پژوهش بیانگر این است که امکان کشف فرار مالیاتی مؤدیان اشخاص حقوقی از طریق به کار گیری روش‌های رگرسیون خطی ساده و الگوریتم درخت تصمیم در نظام مالیاتی ایران وجود دارد. شاخص‌های آماری نشان می‌دهد که دقت و کارایی الگوریتم درخت تصمیم بیشتر از رگرسیون خطی در کشف فرار مالیاتی اشخاص حقوقی بوده است (همانند پژوهش قاسمی و همکاران، ۱۴۰۱؛ نمازی و صادق زاده، ۱۳۹۷؛ کرکوس و همکاران، ۲۰۰۷).

نتایج این پژوهش در بخش تجزیه و تحلیل داده‌ها نشان می‌دهد که الگوی پیش‌بینی فرار مالیاتی اشخاص حقوقی با استفاده از الگوریتم درخت تصمیم عملکرد مناسب را دارد و نتایج این پژوهش اعتبار لازم جهت تصمیم گیری مدیران امور مالیاتی استان در مورد ایجاد مکانیزم اجرایی کنترل فرار مالیاتی را دارد. می‌تواند فرار مالیاتی مؤدیان را کنترل کند، اگرچه در تشخیص مالیات مؤدیان توسط ماموران مالیاتی از این مدل‌ها کمتر استفاده می‌شود. پژوهش حاضر به لحاظ مقایسه‌ای از دو جنبه قابل بحث است:

(اول) اینکه یافته‌ها و دستاوردنها ای پژوهش حاضر، الگوی نسبتاً جامع و کامل بر اساس ویژگی‌های سیستم مالیاتی کشور است و از سایر الگوهای ارائه شده توسعه صاحب‌نظران با توجه به اینکه برای جوامع غیرمالیاتی تدوین شده‌اند؛ مناسب تر است.

(دوم) این الگو به دست آمده از نظر ابعاد، مؤلفه‌ها و شاخص‌های به دست آمده، جامعیت بیشتری را نشان می‌دهد. این موضوع با برخی از پژوهش‌های داخلی و خارجی مقایسه شده است که ضمن برخورداری الگوی پژوهش حاضر از اکثر مؤلفه‌ها و شاخص‌های آنها، تعدادی مؤلفه و شاخص جدید شناسایی و احصاء نموده است که نقطه قوتی برای پژوهش حاضر است.

۷. پیشنهادها

در خصوص پیشنهادهای حاصل از پژوهش به سازمان امور مالیاتی، پیشنهاد می‌شود نتایج حاصل از تحقیق را مدنظر قرار داده و نسبت به برنامه‌ریزی لازم در خصوص به کارگیری تکنیک‌های نوین و کامپیوتربی با توجه به محدودیت منابع انسانی در در دسترس اقدام نماید. علاوه بر این فرار مالیاتی را از طریق تکنیک‌های مورد استفاده در این پژوهش کشف و کنترل نماید؛ ضمن اینکه مؤلفه‌های مورد نظر در این پژوهش که بر فرار مالیاتی اشخاص حقوقی مؤثر بوده‌اند را مورد واکاوی و تجزیه و تحلیل قرارداده و نسبت به برنامه ریزی لازم جهت کنترل آن اقدام نماید. در پژوهش‌های آتی با محوریت این پژوهش، جامعه آماری نمونه برای گروه‌های از صاحبان مشاغل اشخاص حقیقی براساس مشاغل خاص یا گروه‌های از مشاغل پر ریسک که براساس ماده ۹۵ قانون مالیات‌های مستقیم مکلف به ارائه دفاتر و یا استناد و مدارک مالی حسب مورد می‌باشند گسترش یابد و حتی به صورت هم زمان برای اشخاص حقوقی و حقیقی بر اساس نوع صنعت نیز توسعه یابد تا الگوی ارزیابی فرار مالیاتی به صورت جامع تری آزمون شود. این رویکرد می‌تواند به نحو بهتری اثر شاخص‌های همه سطوح و بالاخص سطح کلان اقتصادی را آزمون کند که در این نمونه به عنوان متغیرهای اثرگذار شناسایی نشده‌اند. همچنین می‌توان به موضوع میزان تمکین مالیاتی با استفاده از روش‌های انتخاب ویژگی و رده‌بندی برای تعیین میزان تمکین مؤیدان مالیاتی کشور با به کارگیری از تکنیک‌های مختلف انتخاب ویژگی و سپس از تکنیک‌های مختلف رده‌بندی مانند درخت تصمیم، پرسپترون چند لایه و شبکه‌های بیزین استفاده نمود و در این راستا می‌توان به پژوهش ساتو و همکاران (۲۰۲۱) مراجعه کرد.

با توجه به اینکه پرونده‌های مالیاتی تقریباً با میانگین دو سال تاخیر از تاریخ صورت‌های مالی قطعی می‌شوند و در زمان انجام این تحقیق تقریباً عمدۀ پرونده‌های عملکرد ۱۳۹۸ اطلاعات مالیات قطعی آن در دسترس بود، امکان بسط دوره نمونه به بعد از سال ۱۳۹۸ امکان پذیر نبوده و سال ۱۳۹۸ به عنوان سال پایانی منظور شد. این مسئله از جمله از محدودیت‌های این تحقیق به شمار می‌رود و انتظار می‌رود با افزایش سرعت در رسیدگی و دادرسی پرونده‌های مالیاتی فرایند قطعی سازی زودتر

انجام تا وصول مالیات با اصول انجام و همچنین این محدودیت‌های آماری در انجام چنین تحقیقاتی برطرف شود.

همچنین به سایر پژوهشگران پیشنهاد می‌گردد با توجه به کلاس‌بنای منتخب در این پژوهش، داده‌های سایر استان‌های کشور را مورد تجزیه و تحلیل قرار داده و نتایج را جهت بررسی جامعیت نتایج مقایسه نمایند.

منابع

- امیدی‌پور، رضا؛ پژویان، جمشید؛ محمدی، تیمور و عباس معمارتزاد (۱۳۹۸). «برآورد حجم اقتصاد زیرزمینی و فرار مالیاتی : تحلیل تجربی در ایران». *پژوهشنامه مالیات*، ۲۳(۲۸).
- بوزیده، فرج؛ ابراهیمی درده سجاد (۱۳۹۸). «رتبه بندی عوامل رفتاری مؤثر بر عدم تمکین مالیات بر درآمد مشاغل از منظر خبرگان». *پژوهشنامه مالیات*. ۲۷(۴۲)، صص ۷۲-۴۳.
- ريع بیگی، حامد؛ محمدی پور، رحمت الله؛ نوروش، ایرج و رضا صالحی (۱۴۰۲). «ارائه الگوی روایه مالیاتی با رویکرد تئوری زمینه بنیان». *اقتصاد مالی*، ۶۳(۱۷)، صص ۲۲۶-۱۹۳.
- ساری، محمدعلی؛ اعتمادی، حسین و سحر سپاسی (۱۳۹۶). «ارزیابی ریسک سیاست‌های کمینه‌سازی مالیات شرکت‌ها»، *مطالعات تجربی حسابداری مالی*، ۱۴(۵۴).
- صادقی، حسین و علیرضا شکیبا (۱۳۸۰). «فرار مالیاتی و اندازه اقتصاد زیرزمینی ایران (با روش اقتصادستنجدی فازی)»، *نامه مفید*، دوره ۷ شماره ۲۷ (اقتصاد)، صص ۷۷-۵۵.
- صدمی، علی حسین و رضیه تابنده (۱۳۹۲). «فرار مالیاتی در ایران (بررسی علل و آثار و برآورد میزان آن)»، *پژوهشنامه مالیات*. ۲۱(۱۹)، صص ۱۰۶-۷۷.
- صیام، هدی؛ امین رشتی، نارسیس؛ محراجیان، آزاده و رؤیا سیفی پور (۱۴۰۲). «بررسی عوامل مؤثر بر فرار مالیاتی در شرکت‌های سخت‌مالیات ده با استفاده از روش اثربگذاری غیرخطی»، *اقتصاد کاربردی*، ۱۳(۴۴)، صص ۱۹-۳.
- عبدالله میلانی، مهنوش و نرگس اکبرپور روش (۱۳۹۱). «فرار مالیاتی ناشی از اقتصاد غیررسمی در ایران»، *پژوهشنامه مالیات*، ۲۰(۱۳)، صص ۱۶۸-۱۴۱.

- عسکری، علی (۱۳۹۲). «کالبد شکافی اصلاح نظام مالیاتی کشور: چالش‌ها، برنامه‌ها و عملکرد نظام مالیاتی»، فصلنامه سیاست‌های مالی و اقتصادی، ۱(۲)، صص ۸۵-۱۲۰.
- فلاح شمس، میرفیض؛ زمودیان، غلامرضا؛ انواری رستمی، علی‌اصغر و هوشنگ خرم‌نیا (۱۴۰۱). «مقایسه مدل شبکه عصبی و رگرسیون لاجیت در اعتبارسنجی مؤیدیان مالیاتی»، پژوهشنامه مالیات، ۵۵(۳۰)، صص ۱۰۸-۱۲۶.
- قاسمی، محمد؛ عابدی، صادق و علی محتشمی (۱۴۰۱). «ارائه الگویی پیش‌بینی فرار مالیاتی اصناف مبتنی بر تکنیک داده کاوی»، پژوهشنامه مالیات، ۳۰(۵۶)، صص ۷-۳۰.
- مسیحی، محمد؛ یعقوب نژاد، احمد؛ کیقبادی، امیررضا و تقی توابی (۱۳۹۸). «استفاده از تکنیک‌های داده کاوی برای سنجش ریسک مالیاتی مؤیدیان مالیات بر ارزش افزوده»، دانش سرمایه‌گذاری، ۳۲(۸)، صص ۳۴۷-۳۶۳.
- نجفی، سمیه؛ صالحی، الله‌کرم و هوشنگ امیری (۱۴۰۱). «ارائه الگویی برای کشف تقلب مالیاتی بر اساس تیپ‌های شخصیتی مدیران مالی شرکت‌ها با استفاده از رویکرد شبکه‌های عصبی»، پژوهشنامه مالیات، ۳۰(۵۳)، صص ۷۱-۹۶.
- نمازی، محمد و محمد صادق‌زاده مهارلوئی (۱۳۹۶). «پیش‌بینی فرار مالیاتی با استفاده از الگوریتم‌های داده کاوی درخت تصمیم»، حسابداری مالی، ۹(۳۶)، صص ۷۶-۱۰۱.
- نسل موسوی، سیدحسین؛ حسینی شیروانی، میرسعید و محمود نظرپور (۱۳۹۹). «ارائه مدل پیش‌بینی فرار مالیاتی بر مبنای الگوریتم درخت تصمیم ID ۳ و شبکه بیزین». پژوهشنامه مالیات، ۲۸(۴۵)، صص ۵۹-۸۷.

- Abdixhiku L.B. and G. Krasniqi** (2017). “Pugh and I. Hashi, Firm -level Determinants of Tax Evasion in Transition Economies”, Economic Systems. 41(3), pp. 354 -366
- Agrawal P., Hattan F.A., Talari Ga. and W. M. Ali** (2021). “Metaheuristic Algorithms on Feature Selection: A Survey of one decade of research (2009-2019)”. IEEE Access 9 (2021), pp. 26766-26791.
- Annan B., Bekoe W. and E. Nketiah-Amponsah (2013). “Determinants of tax evasion in Ghana: 1970-2010”. *International Journal of Economic Sciences and Applied Research*, 6(3).

- Djajanti A.** (2020). "Developing the Voluntary Tax Payer Compliance: the Scale of the Tax Authority's Power, Trust and the Fairness of the Tax System", *Indonesian Journal of Business and Entrepreneurship*, Vol. 6, No. 1.
- Didimo W., Grilli L., Liotta G., Menconi L., Montecchiani F. and D. Pagliuca** (2020). *Combining Network Visualization and Data Mining for Tax Risk Assessment*. IEEE Access, 8, 16073-16086.
- Di Gioacchino D. and D. Fichera** (2020). "Tax Evasion and Tax Morale: A Social Network Analysis". *European Journal of Political Economy*. No. 65, pp. 121 -149.
- Giovanni D. De, Lamantia F. and M. Pezzino** (2019). "A Behavioral Model of Evolutionary Dynamics and Optimal Regulation of Tax Evasion", *Structural Change and Economic Dynamics*, No. 50, pp. 79–89.
- Hafez A., Ibrahim H.M., Zawbaa E.E. and Aboul Ella Hassanien** (2016). "Sine Cosine Optimization Algorithm for Feature Selection. In 2016 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)", pp. 1-5.
- Kenno B.G.** (2020). "Factors Affecting Perception of Taxpayers towards the Seriousness of Tax Evasion in Bale Robe Town Adminidstration, Oromia Ethiopia". *International Journal of Finance and Accounting*, No. 2, pp. 21-30.
- López C.P., Rodríguez M.J.D. and S.L. Santos** (2019). "Tax Fraud Detectio through Neural Networks: An Application Using a Sample of Personal Income Taxpayers", *Future Internet*, No. 11, pp. 1 -13.
- Nguyen T.H.** (2022). "The Impact of Non-economic Factors on Voluntary Tax Compliance Behavior: a Case Study of Small and Medium Enterprises in Vietnam". *Economies*, 10(8), pp. 179.
- Ruan J., Yan Z., Dong B., Zheng Q. and B. Qian** (2019). "Identifying suspicious groups of affiliated-transaction-based tax evasion in big data". *Information Sciences*, No.477, pp.508-532.
- Ravisankar P., Ravi V., Rao R. and I. Bose** (2011). "Detection of Financial Statement Fraud and Feature Selection Using data Mining Techniques". *Decision Support Systems*, 50(2), pp. 491–500.
- Sneha N. and Tarun Gangil** (2019). "Analysis of diabetes mellitus for early prediction using optimal features selection". *Journal of Big data* 6, No. 1, pp. 1-19.
- Vanhoeyveld J., David M. and Bruno Peeters.** (2020). "Value-added tax Fraud Detection with Scalable Anomaly Detection Techniques". *Applied Soft Computing*, Volume 86, January 2020.
- Yamen A., Allamc A., Bani A. and M.A. Uyard** (2021). "Impact of Institutional Environment Quality on Tax Evasion: A Comparative Investigation of Old Versus New EU members", *Journal of International Accounting, Auditing and Taxation*, No.42, pp. 100-352.