نشریـه علمـی حسابداری مدیریت
سال سیزدهم/ شماره چهل‌وششم/ پائیز ۱۳۹۹

ارائه مدل شناسایی تقلب مالیاتی بر مبنای ترکیب الگوریتم درخت تصمیم ID3 بهبود یافته و شبکه های عصبی پرسپترون چندلایه

اکبر جوادیان کوتنائی[۱]
عباسعلی پورآقاجان سرحمامی[۲]
میرسعید حسینی شیروانی[۳]

تاریخ دریافت: ۹۸/۰۷/۰۷ تاریخ پذیرش: ۹۸/۱۰/۰۸

چکیده

درآمدهای مالیاتی یکی از مهم‌ترین منابع درآمدی دولت و تأمین‌کننده بخش عمده‌ای از هزینه‌های دولـت اسـت. در سالهای اخیر تقلب در صورت‌های مالی و اظهارنامه های مالیاتی به طور فزاینده‌ای به یک مشکل جدی برای کسب و کار، دولت و سرمایه‌گذاران تبدیل شده است. اکثر مؤدیان مالیاتی به دنبال راهی برای دستکاری در صورتهای مالی و کاهش سود مشمول مالیات ابرازی خود می‌باشند. از این‌رو، شناسایی متقلبین مالیاتی و شرکتهایی کـه بـه تقلـب در صورتهای مالی می‌پردازند به امری حیاتی برای دولت تبدیل شده است.

هدف از این تحقیق ارائه مدلی است که در آن از الگوریتم درخت تصمیم‌گیریID3 بهبود یافته استفاده شده است. همچنین برای بهبود عملکرد و دقت آن، با شبکه های عصبی پرسپترون چندلایه بهینه سازی شده توسط الگوریتم ژنتیک ترکیب گردید تا نسبت‌های مالی مرتبط با تقلب مالیاتی انتخاب نموده و سربار محاسباتی کاهش یابد. درختی که در مدل پیشنهادی ایجاد می‌شود دارای کمترین عمق ممکن می باشد که از این‌رو دارای سرعت بالا و سربار محاسباتی پایینی می‌باشد. بدین منظور صورتهای مالی 60 شرکت پذیرفته شده در بورس اوراق بهادار تهران در سالهای ۱۳۹۴ لغایت ۱۳۹۶ بررسی و ۵۴ نسبت مالی از آن‌ها استخراج گردید که به وسیله آزمون ANOVA تعداد ۲۳ نسبت و نهایتاً توسط شبکه های عصبی تعداد ۷ نسبت مرتبط با تقلب مالیاتی، به عنوان داده های ورودی مدل انتخاب گردید. مدل ارائه شده با دقت 81/4 درصد، در شناسایی شرکتهای دارای تقلب مالیاتی، موفق بوده که نسبت به الگوریتم آدابوست دارای بالاترین دقت و قدرت پیش بینی بوده است.

واژه‌های کلیدی: تقلب مالیاتی، مالیات ابرازی، درخت تصمیم بهبود یافته، الگوریتم ژنتیک، الگوریتم آدابوست.

۱- دانشجوی دکتری حسابداری، گروه حسابداری، واحد قائم شهر، دانشگاه آزاد اسلامی، قائم شهر، ایران.
۲- استادیار، گروه حسابداری، واحد قائم شهر، دانشگاه آزاد اسلامی، قائم شهر، ایران. (نویسنده مسئول) abbas_acc46@yahoo.com
۳- استادیار، گروه مهندسی کامپیوتر، واحد ساری، دانشگاه آزاد اسلامی، ساری، ایران

۱- مقدمه

یکی از منابع دولت، وصول مالیات از شرکتهاست. مبانی نظری و شواهد تجربی نشان می‌دهد شرکتها درتلاش‌اند تا مالیات بردرآمد خودرا کاهش داده یا به تعویق بیندازند. تقلب مالیاتی، اجتناب مالیاتی، فرارمالیاتی، مدیریت سود و غیره، ابزارهایی هستندکه شرکت‌ها ممکن است برای کاهش مالیات درصورت لزوم، با توجه به شرایط و موقعیت، از آنها استفاده کنند. تقلب مالیاتی نوعی تخلف قانونی است، اما اجتناب از مالیات، در واقع نوعی استفاده ازخلأهای قانونی در قوانین مالیاتی برای کاهش مالیات است. بنابراین، ازآنجا که اجتناب مالیاتی فعالیتی به ظاهر قانونی است، به نظر می‌رسدکه بیشتر از تقلب مالیاتی در معرض دید باشد. (عرب صالحی و هاشمی، ۱۳۹۴).

در محیطی فعال از تقلب، مکانیزم‌های کشف تقلب با کمک رایانه بسیار مؤثرتر و کاراتر خواهد بود. تکنولوژی‌های مبتنی بر آمار و یادگیری ماشینی یک راهکار اثربخش برای پیشگیری و کشف تقلب هستند اما مرتکبین تقلب خود را انطباق می‌دهند و به‌طورمعمول قادر به کشف راه‌هایی برای دور زدن این تکنولوژی‌ها هستند. تکنیک‌های کنونی کشف تقلب در اکثر موقعیت‌های مرتبط با تقلب از اصول داده‌کاوی مشابهی استفاده می‌کنند اما می‌توانند از لحاظ دانش قلمرو تخصص متفاوت باشند. زمانی که مدیران مالی درگیر در تقلب از نرم‌افزارها و تکنیک‌های کشف تقلب اطلاع کافی دارند، روش‌هایی را برای ارتکاب تقلب به‌کار می‌گیرند که کشف آنها به‌ویژه با استفاده از تکنیک‌های فعلی دشوار است. ازاین‌رو، یک نیاز مبرم برای روش‌هایی که نه‌تنها کارآ باشند بلکه برای کشف فریبکاری‌های مالیاتی انطباقی و نوظهور مؤثر هستند، وجود دارد.

در حالت پیچیده‌تر، هر ساله داده‌های بسیاری تولید می‌شوند و کشف اطلاعات متقلبانه مستلزم تکنیک‌های کاراتری جهت کاوش داده‌ها می‌باشد. در راستای شناسایی این معضل اکثریت روش های ارائه شده مبتنی بر الگوریتم‌های موجود می‌باشند و تنها سعی شده است تا از روش‌های انسانی و یا داده کاوی-های ساده شناسایی صورت گیرد که این کار خود سربار بالایی دارد و همچنین دارای هزینه بالایی نیز می‌باشند. روش‌های داده کاوی که تا کنون برای این کار ارائه شدند دارای سربار بالای محاسباتی و یا دقت پایین بودند.

هدف از این تحقیق ارائه مدلی جهت شناسایی تقلب مالیاتی است که در آن از درخت تصمیم گیری ID3 بهبود یافته به همراه ماشین بردار پشتیبان به عنوان یک روش ترکیبی استفاده شده است و نیز در کنار آن برای بهبود عملکرد و دقت از الگوریتم ژنتیک و شبکه های عصبی پرسپترون چندلایه در جهت انتخاب ویژگی های موثرتر بهره گرفته شده است تا سربار محاسباتی کاهش یابد. درختی که در روش پیشنهادی ایجاد می شود دارای کمترین عمق ممکن می باشد که از این‌رو دارای سرعت بالا و سربار محاسباتی پایینی می‌باشد.

در ادامه این پژوهش به تشریح مفهوم تقلب مالیاتی و الگوریتم‌های مورد استفاده پرداخته می‌شود و در نهایت با ارائه مدل و ارزیابی آن، به مقایسه مدل ارائه شده با الگوریتم آدابوست[1] پرداخته و قدرت و دقت پیش‌بینی هر یک بررسی می‌گردد.

۲- مفهوم تقلب مالیاتی

به عقیده واتس (۲۰۰۳) از آنجا که گزارشگری مالی مقدمه پرداخت مالیات می‌باشد پس ملاحظات گزارشگری بر ملاحظات مالیاتی اولویت دارد. هدف اولیۀگزارشگری مالی فراهم آوردن اطلاعات اقتصادی مفید برای استفاده کنندگان بیرونی است، در حالی که هدف اولیه از ارایه اطلاعات حسابداری مالیاتی انعکاس اطلاعات به گونه ای است که به پرداخت حداقل مالیات منجر گردد. هرچند دستورالعمل های مالیاتی با دستورالعمل‌های اصول پذیرفته شده حسابداری به گونه‌ای پیوند خورده‌اند که واحدهای تجاری را قادر ساخته‌اند تا با استفاده از رویه هایی، دسترسی به هر دو هدف را تا حدودی محقق سازد، اما این اختلاف اهداف طبق اصول پذیرفته شده حسابداری و اصول مالیاتی به اختلاف بین سود قبل از مالیات انعکاسی در

صورت سود و زیان و درآمد مشمول مالیات در اظهارنامه مالیاتی منجر می گردد؛ پس واحد تجاری برای تحمل مالیات کمتر سعی می‌کنند درآمد مشمول مالیات کمتری را نشان دهد.

موضوع مالیات برعملکرد، بخش مهمی از حوزۀ ادبیات مالی و اقتصادی هر کشور محسوب می‌شود. محیط اطلاعاتی حسابداری در ایران به گونه‌ای است که مالیات یکی از ارکان اثرگذار بر گزارشگری مالی به شمار می‌رود. این موضوع به اندازه ای اهمیت دارد که طبق مفاد استاندارد حسابداری شمارۀ ۲ ایران با عنوان "صورت جریان وجوه نقد"، برای گزارشگری جریان های نقدی مختص به مالیات بر عملکرد طبقه‌بندی جداگانه‌ای پیش‌بینی شده است. علاوه براین، بنا بر الزامات افشای استانداردهای حسابداری ایران، در یادداشت‌های توضیحی صورت‌های مالی باید اطلاعات تفصیلی دربارۀ مالیات ابرازی، تشخیصی و قطعی افشا شود (استانداردهای حسابداری ایران، ۱۳۹۳)؛ ضمن آنکه عمدۀ منبع درآمدهای مالیاتی در ایران به مالیات بردرآمد (عملکرد) شرکتها اختصاص دارد (عرب مازار ودهقانی، ۱۳۸۸). در این راستا، پرداخت مالیات نیز از اهمیت بسزایی برخورداراست؛ به‌طوریکه همیشه دولت‌ها درصددند با وضع قوانینی زمینه‌های اینکار را فراهم کنند، اما مؤدیان مالیاتی بنا به دلایل گوناگون از راه‌های تقلب مالیاتی، اجتناب مالیاتی یا فرار مالیاتی، سعی در پرداخت مالیات کمتری دارند. نمود این موضوع در برخی از متغیرهای مالی مشاهده می شود. دراین راستا می‌توان به تفاوت بین سود حسابداری و سود مشمول مالیات و همچنین نرخ مؤثر مالیات به عنوان معیاری از وضعیت اجتناب مالیاتی اشاره کرد. (پورحیدری و گل محمدی ۱۳۹۴).

انگیزه مالیاتی شرکتها و واکنش مدیران به آن بسیار جذاب است، مخصوصاً در این برهه از زمان، که مشاهده می‌نماییم این تردید و گمان قوت گرفته است که شرکت‌های بزرگ، مالیات کمتر از واقع پرداخت می‌نماید. برخی از مطالعات نشان می‌دهند در سال‌های اخیر تصور اینکه شرکت‌های بزرگ مالیات واقعی خود را نمی‌پردازند شدت یافته است؛ این تصور بعد از رسوایی‌های اخیر تقویت گردیده است. انرون نمونه خوبی از این نوع شرکت‌ها است؛ این شرکت تا قبل از ورشکستگی خود مالیات ناچیزی پرداخت می‌کرد، درحالیکه در همان سال‌ها درآمدهای هنگفتی شناسایی و گزارش می‌نمود. در واقع انرون با استفاده از ضعف‌های قانونی و استانداردی و همچنین با بکارگیری حسابداری متهورانه و تقلب در گزارشگری مالی نشان داد چگونه می‌توان چندین سال بازار، دولت و جامعه را فریب داد و در عین حالی که درآمدهای هنگفتی شناسایی نمود از شناسایی و پرداخت مالیات اجتناب نمود. سلب اعتماد عمومی که رسوایی‌ها بوجود آوردند زنگ خطری را به صدا درآوردند تا افراد و گروه‌های مختلفی به این مسائل بیندیشند و درصدد راهکارهایی باشند تا از وقوع مجدد این موارد اجتناب شود. قانون ساربینز آکسلی (۲۰۰۲) و سپس قانون داد فرانک (۲۰۰۹) و تغیراتی که در استانداردها، قوانین، مجازات‌ها و غیره بوجود آمد واکنش نهادها و دولت به این رسوایی‌ها بود. (مهرانی و سیدی، ۱۳۹۳).

با عنایت به موارد مطروحه فوق، در این پژوهش هرگونه فعالیت متهورانه و دستکاری در صورتهای مالی و اظهارنامه مالیاتی به منظور کاهش سود مشمول مالیات، به عنوان تقلب مالیاتی تلقی می‌شود.

۳- الگوریتم ژنتیک[۲]

هدف اصلی الگوریتم ژنتیک انتقال خصوصیات موروثی توسط ژن‌ها است. این الگوریتم یک روش احتمالی است که از اصول علم وراثت و تکامل زیستی استفاده می‌کند. الگوریتم‌های ژنتیک با به‌کارگیری اصل تداوم بهترین ها برای تولید برآورد هر چه بهتر یک جواب (کروموزوم‌ها) روی تعدادی از جواب‌های بالقوه عمل می کند. هر پاسخ ممکن برای مسئله مطرح شده را یک کروموزم و مجموعه‌ای از کروموزم‌ها را جمعیت می‌نامند (نسل موسوی، ۲۰۱۱).

خصوصیات یک الگوریتم ژنتیک به شرح زیر است:

الگوریتم ژنتیک تنها به یک نکته توجه ندارد، بلکه نکات بسیاری را در فضای پژوهش، برای بهینه

ساختن پاسخ به دست آمده، در نظر می گیرد. به پاسخ این الگوریتم تا حد بسیار زیادی می توان اعتماد کرد و پاسخی که می یابد به احتمال زیاد، بهترین پاسخ ممکن است. در این روش از قوانین احتمال به جای قوانین قطعی استفاده می شود. اما مشکل اصلی الگوریتم ژنتیک علی رغم سادگی پیاده سازی، هزینه اجرای آن است. اغلب حل یک مسئله نیازمند تولید چندین هزار نسل از کروموزم ها است و این مسئله نیاز به زمان زیادی دارد و همین امر، گاهی استفاده از الگوریتم را با مشکل مواجه می کند (شین و لی، ۲۰۰۲).

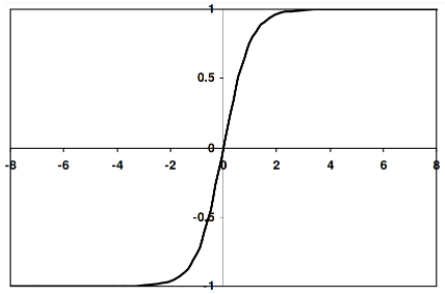
۴- شبکه‌های عصبی مصنوعی[۳]

امروزه کاربردهای بسیاری برای شبکه‌های عصبی بوجود آمده است و با توجه به نوع و کاربرد ساختار و قوانین یادگیری در شبکه می‌تواند متفاوت باشد. بطور خلاصه مهمترین شبکه‌های عصبی را می‌توان به صورت زیر تقسیم بندی نمود:

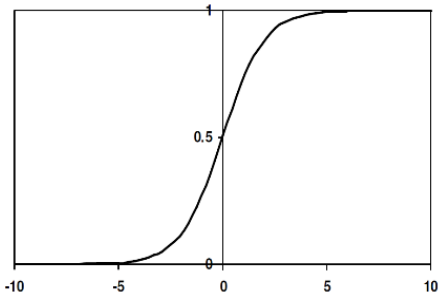
پرسپترون چندلایه (MLP)[۴]، شبکه خود سازمانده (SOM)[۵]، LVQ[۶]، هاپفیلد[۷]، تأخیر زمانی (TDNN)[۸] و RBF[۹].

در این پژوهش از شبکه عصبی پرسپترون چندلایه استفاده شده است. شبکه MLP از چند لایه تشکیل شده است. لایه ورودی، لایه خروجی و لایه یا لایه‌های مخفی که خروجی لایه اول، بردار ورودی لایه دوم به حساب می‌آید. به همین ترتیب خروجی لایه دوم، بردار ورودی لایه سوم را تشکیل می‌دهد. خروجی‌های لایه دوم پاسخ واقعی شبکه را نشان می‌دهند.

نرون‌های موجود در لایه بالادست به نرون‌های موجود در لایه پائین دست ارتباط دارند. نقش هر نرون محاسبه مجموع وزن داده شده پترون ورودی (Net) و سپس گذراندن این مجموع از یک تابع به نام تابع انتقال می‌باشد. تابع انتقال می‌تواند یک تابع خطی یا غیرخطی باشد. دو نوع از توابع مرسوم در شبکه پرسپترون چند لایه، تابع سیگموید و تانژانت سیگموید می‌باشد که در شکل‌های (۱) و (۲) این توابع نشان داده شده‌اند. (Vakil-Baghmishen,2002).



شکل (۱): تابع تانژانت هایپر پولیک

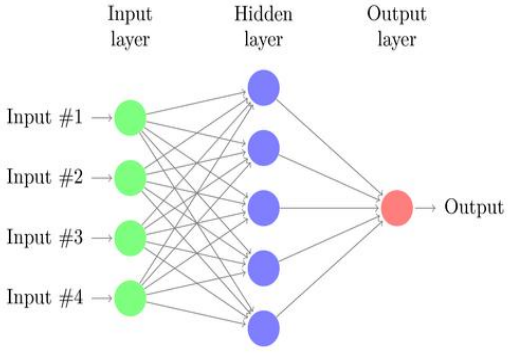


شکل (۲): تابع سیگموید

نحوه عمل پرسپترون چندلایه بدین صورت است که الگویی به شبکه عرضه می‌شود و خروجی آن محاسبه می‌گردد. مقایسه خروجی واقعی و خروجی مطلوب، باعث می‌شود که ضریب وزنی شبکه تغییر یابد به طوریکه در دفعات بعد خروجی درست‌تری حاصل می‌شود. قاعده فراگیری میزان کردن، ضرایب وزنی شبکه را بیان می‌کند. (صدر موسوی و رحیمی، ۱۳۸۸)

۵- درخت تصمیم[۱۰]

درخت تصمیم گیری آرایه‌ای از پارامترها را به عنوان ورودی دریافت می کند و نهایتاً یک تصمیم را برمی‌گرداند که در حقیقت مقدار خروجی تخمین زده شده به ازای ورودی می‌باشد. درخت تصمیم گیری (شکل شماره 3)، با انجام رشته‌ای از تست‌ها بر روی مقادیر ورودی مقدار تخمینی برای خروجی را تعیین می‌کند.



شکل (۳): نمونه‌ای از درخت تصمیم

هرگره درونی در درخت تصمیم گیری، بیانگر یکی از پارامترهای ورودی است و شاخه‌های خارج شده از این گره هر کدام بیانگر یک مقدار یا یک دسته مقادیر از آن پارامتر می‌باشد. هر برگ درخت بیانگر مقداری است که اگر به آن برگ برسیم توسط درخت تصمیم‌گیری برگردانده می‌شود.

نحوه یادگیری در سیستم درخت‌های تصمیم‌گیری به طور خلاصه بدین صورت خواهد بود که با استفاده از نمونه های آموزشی و روابط تئوری اطلاعات[11]، ورودی‌ها را بر حسب درجه اهمیت آن‌ها در تعیین مقدار خروجی مرتب می کنند. سپس درخت تصمیم‌گیری را طوری خواهند ساخت که ورودی های مهم‌تردر سطوح بالاتر درخت مورد آزمایش قرار گیرند و بالعکس.

درخت‌های تصمیم‌گیری یکی از ساده‌ترین و درعین حال موفق‌ترین روش‌های یادگیری ماشین می باشند. همان طور که مشهود است این درخت ها پاسخ نمونه‌های مجهول را به سرعت تخمین می‌زنند و پروسه یادگیری در آنها بسته به تعداد و نوع ورودی‌ها ممکن است بسیار زمانبر باشد. (همراز، ۱۳۸۵)

۶- الگوریتم آدابوست

آدابوست یکی از مشهورترین روش های الگوریتم یادگیری است. مهمترین ویژگی این روش پیاده سازی ساده آن است. در این روش، طبقه کننده های ضعیف تر با هم ترکیب شده تا مدلی با قدرت پیش بینی بالا را تشکیل دهند (هو و دیگران، ۲۰۰۸).

آدابوست یک روش است که از ویژگی طبقه بندی کردن به وسیله بهبود در صحت و درستی طبقه هایش حداکثر استفاده را می کند. اگر چه چندین نسخه از الگوریتم شبیه به آدابوست وجود داشت اما نسخه آدابوست که بوسیله فراند و اسکیپر تهیه شد به دلیل ساده بودن معروف گردید در این روش تنها دو طبقه وجود دارد یک طبقه از گروه آموزشی که به وسیله معادله زیر مشخص می شود:

$$Tn=[(X_1, Y_1),(X_2 ,Y_2),.... (X_n\ Y_n)$$

مقدار Y بین [-1,1] است، وزن اختصاص یافته به هر مشاهده X_i گروه اولیه برای 1/n است این ارزش بعد از هر مرحله به روز می شود:

$$\int(x) = \sum_{t=1}^{T} \alpha_t\, h_t(x)$$

$$ht(x): X \rightarrow \{-1, +1\}$$

۷- پیشینه تحقیق

همتی و همکاران (۱۳۹۸) در پژوهشی به بررسی رابطه میان نوع اظهارنظر حسابرس و اندازه موسسه حسابرسی با اجتناب مالیاتی پرداختند. بدین ترتیب ۷۱ شرکت طی ۱۳۸۴ تا ۱۳۹۴ در بورس اوراق بهادار تهران مورد بررسی گرفته و نتایج حاکی از آن بود که اظهارنظر حسابرس و اندازه موسسه حسابرسی با اجتناب مالیاتی شرکت رابطه معناداری دارند. به عبارت دیگر، شرکتهایی که به دلیل عدم رعایت قوانین و مقررات از حسابرس خود اظهار نظر غیر مقبول دریافت نموده‌اند، احتمال اجتناب مالیاتی در آنها افزایش می‌یابد.

تشدیدی و همکاران (۱۳۹۸) در پژوهش خود با عنوان ارائه رویکردی نوین در پیش‌بینی و کشف تقلب صورت‌های مالی با استفاده از الگوریتم زنبور عسل به مطالعه بر روی ۱۲۰ شرکت در دوره زمانی ۱۳۸۵ لغایت ۱۳۹۶ در بورس اوراق بهادار تهران پرداختند و به این نتیجه رسیدند که الگوریتم زنبور عسل با دقت ۸۲,۵ نسبت به الگوریتم ژنتیک با ۷۷,۵ درصد و رگرسیون لجستیک با دقت ۷۲,۵ درصد دارای

عملکرد بهتری در شناسایی شرکتهای مشکوک به تقلب در صورتهای مالی می باشد.

موسایی و دیگران (۱۳۹۸) در پژوهشی به تدوین مدل کشف تقلب با استفاده از رویکرد ترکیبی برپایه مدل تحلیل عاملی و روش شبکه عصبی مصنوعی در بازه زمانی ۱۳۹۲ تا ۱۳۹۳ در ۱۴۰ شرکت فعال در بورس اوراق بهادار تهران پرداختند. برای دسته بندی شرکت‌ها با احتمال گزارشگری متقلبانه وغیر متقلبانه از مدل نمرهM بنیش استفاده شده است که ۷۸ شرکت دارای احتمال گزارشگری متقلبانه و۶۲ شرکت دارای احتمال گزارشگری غیرمتقلبانه بوده است. نتایج نشان داد که ساختار گزارش شده مدل شبکه عصبی که دارای ۷ نرون در لایه پنهان است از دقت و عملکرد بالاتری نسبت به سایر ساختارهای بررسی شده بوده است. نتایج حاکی است که دقت دسته-بندی شرکت-های متقلب و شرکت-های غیرمتقلب و عملکرد کلی در روش شبکه عصبی مصنوعی به ترتیب۵۷,۶۹ ٪و۷۲,۷۳ ٪ و۶۲,۱۶ ٪ بوده است.

نمازی و صادق زاده مهارلوئی نیز (۱۳۹۷) در پژوهش خود با عنوان پیش بینی فرار مالیاتی با استفاده از الگوریتم‌های داده کاوی درخت تصمیم با بررسی ۱۰۸۱ سال-شرکت نتیجه گرفتند که روشهای جنگل تصادفی، کاهش خطای هرس، ریشه تصمیم و درخت تصادفی از دقت و کارایی بیشتری در پیش‌بینی فرار مالیاتی ، J48،LMT برخوردار هستند.

تقوی فرد و همکاران (۱۳۹۶) در پژوهشی به بررسی تحلیل آینده نگر تشخیص فرار مالیاتی مودیان مالیات بر ارزش افزوده در دوره زمانی ۱۳۸۸ تا ۱۳۹۳ با استفاده از الگوریتم‌های Naive Bayes ، درخت تصمیم، KNN، k-medoids و k-means پرداخته‌اند . نتایج پژوهش نشان داد که از میان الگوریتم‌های مذکور، روش درخت تصمیم با خطای ۰,۰۵ و اعتبارسنجی ۰,۱۷ نتیجه بهتری کسب کرد.

وانهیویلد و دیگران (۲۰۲۰) در پژوهش خود تحت عنوان کشف تقلب مالیات بر ارزش افزوده با تکنیکهای تشخیص ناهنجاری بدون نظارت به تشخیص تقلب مالیات بر ارزش افزوده در بلژیک به وسیله تکنیکهای بهینه تشخیص ناهنجاری بدون ناظر پرداختند. بدین ترتیب آنها نشان دادند، تشخیص ناهنجاری به وسیله الگوریتم های بدون نظارت، قدرت پیش بینی کننده بالایی برای تشخیص متقلبین مالیات بر ارزش افزوده دارد. همچنین شاخص های جدید تقلب مالیات بر ارزش افزوده ارائه شده است که تقلب را با موفقیت تشخیص می‌دهد. در نهایت الگوریتم های توسعه یافته جهت تشخیص سریع ناهنجاری و مقیاس پذیری آنها را معرفی نمودند.

دیدیمو و دیگران (۲۰۱۸) برای کشف فرار مالیاتی به توصیف سیستم پشتیبانی تصمیم بر مبنای زبان دیداری و تکنیکهای پیشرفته مصورسازی شبکه پرداختند. به وسیله این سیستم، این امکان وجود دارد که گرافهای زیرمجموعه نمونه‌های مشکوک را با توجه به تطابق با الگوهای موجود رسم نموده و باعث ادغام نتایج و ایجاد شاخص‌هایی برای طبقه‌بندی کردن مالیات‌دهندگان بر مبنای ریسک مالی شوند. نتایج نشان دهنده مؤثربودن سیستم پیشنهادی بوده است.

لاری دشتبیاز و دیگران (۲۰۱۵)، در تحقیقی تحت عنوان فرآیند جستجو و کشف اطلاعات برای تقلب در صورتهای مالی به یک مرور کلی از فرایندهای داده‌کاوی مورد استفاده برای شناسایی تقلب مالی، به ویژه تقلب در صورتهای مالی شرکتها، پرداختند. آنها در نتایج پژوهش خود اظهار داشتند که مهمترین روش‌های مورد استفاده برای شناسایی تقلب مالی شامل رگرسیون لجستیک، شبکه های عصبی، شبکه های استنتاج بیزی و درخت تصمیم گیری است که راه حل‌های مهمی برای مشکلات ذاتی شناسایی و طبقه بندی داده‌ها هستند.

ووا، اوو، لین، چنگ و ین (۲۰۱۳)، با اعمال یک سیستم مبتنی بر داده کاوی بر روی پایگاه داده مالیات برای ارزش افزوده نتیجه گرفتند که استفاده از سیستم‌های مبتنی بر داده‌کاوی موجب بهبود چشمگیری در تشخیص کسب و کارهای فراری در حوزه مالیات بر ارزش افزوده می‌شود.

راویسانکار، راوی، رائو و بس (۲۰۱۱) مدل‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان، برنامه ریزی ژنتیک، شبکه عصبی گروهی مدیریت داده‌ها، رگرسیون لجستیک و شبکه عصبی احتمالی را بر روی ۲۰۲ شرکت چینی بررسی نمودند تا تقلب را با استفاده و بدون استفاده از انتخاب ویژگی کشف نمایند. در شرایط استفاده از فرآیند انتخاب ویژگی برنامه ریزی ژنتیک و شبکه عصبی احتمالی دقت های بالاتر را نتیجه دادند. در شرایط عدم استفاده از انتخاب ویژگی، شبکه عصبی احتمالی دارای بالاترین دقت بوده است. روش مورد استفاده برای انتخاب ویژگی، روش ساده آماری آزمون تی می‌باشد. میانگین دقت کلی مجموعه مدل‌ها با اعمال اعتبارسنجی ضربدری ۱۰ بخشی بین ۷۰٪ تا ۹۲٪ می‌باشد.

ژو و کاپور (۲۰۱۱)، در پژوهشی با استفاده از مجموعه مدل‌های رگرسیونی، درخت تصمیم، شبکه عصبی و شبکه بیزین اقدام به تشخیص فرار مالیاتی نمودند. همچنین در ادامه پژوهش مدل سطح پاسخ را نیز با مدلهای فوق ترکیب نموده و نتیجه گرفتند که این مدل ترکیبی نتیجه بهتری خواهد داشت.

کِرکوس، اسپاتیس و مانولوپولوس (۲۰۰۷)، با استفاده از درخت تصمیم، شبکه عصبی مصنوعی و شبکه بیزینی و ورودی‌هایی مبتنی بر نسبت‌های مالی اقدام به تشخیص فرار مالیاتی کرده‌اند. براساس نتایج این پژوهش، دقت کلی مدل‌های طبقه بندی فوق، در درخت تصمیم ۷۳/۶٪ ، در شبکه عصبی ۸۰٪ و در شبکه بیزینی ۹۰/۳٪ می‌باشد که به وضوح برتری مدل شبکه عصبی بیزین را به نمایش می‌گذارد.

۸- روش تحقیق و تشریح مدل پیشنهادی

برای تشخیص و طبقه بندی واحدهای اقتصادی به شرکتهای متقلب یا سالم در گزارشگری مالی، چارچوب نظری خاصی وجود ندارد. با توجه به استانداردهای شماره ۲۴۰و ۴۵۰ حسابرسی ایران، می‌توان برمبنای مقدار و محتوای ویژگیهای داده های صورتهای مالی حسابرسی شده توسط حسابداران رسمی یا سازمان حسابرسی، معیارهایی را برای محسوب نمودن تحریف ارائه نمود. لذا با توجه تحقیقات پیشین از قبیل تحقیق همتی و همکاران (۱۳۹۸)، داغمه چی فیروزجایی (۱۳۹۳)، اعتمادی و زلقی (۱۳۸۸)، فرقاندوست حقیقی و بروازی (۱۳۸۸) و مهام و دیگران (۱۳۹۱) جهت طبقه بندی شرکتهای نمونه به متقلب و سالم، می‌بایست معیارهای زیر طی حداقل سه سال (سال ۱۳۹۴ الی ۱۳۹۶) در صورتهای مالی شرکت‌های متقلب، صادق بوده و وجود سه شرط زیر به عنوان شرایط طبقه بندی در گروه شرکتهای دارای احتمال تقلب می باشد. این شروط عبارتند از:

۱- وجود اختلافات مالیاتی با حوزه مالیاتی طبق یادداشت ذخیره مالیات بردرآمد و پرونده مالیاتی و بند شرط گزارش حسابرسی. ۲- وجود تعدیلات سنواتی با اهمیت و صورتهای مالی تجدید ارائه شده.۳- اظهارنظر غیر مقبول حسابرسی

دلایل انتخاب این معیارها این است که در مورد معیار اول، اختلافات مالیاتی بطور عمده ناشی از تفسیر نادرست قوانین مالیاتی و اشتباه سهوی یا عمدی دربکارگیری بندهای قوانین ذیربط و در برخی موارد تاخیر در شناسایی مالیات و حفظ نقدینگی و سایر موارد احتمالی تقلب و فرار مالیاتی است.

در مورد معیار دوم؛ موارد اشتباه و دستکاری اقلام بویژه اقلام سود و زیانی در سنوات قبل زمینه ساز بروز ارائه مجدد صورتهای مالی و دلیلی بر احتمال تقلب درصورتهای مالی است. به موجب بند ۳۸ استاندارد حسابداری شماره ۶، اثر تعدیلات سنواتی باید از طریق اصلاح مانده سود (زیان) انباشته ابتدای دوره در صورتهای مالی منعکس گردد. همچنین به دلیل تأثیر با اهمیت تعدیلات سنواتی بر درآمد مشمول مالیات، در تاریخ ۱۳۹۱/۱۰/۲۵ دستورالعملی تحت عنوان رسیدگی به حساب تعدیلات سنواتی به شماره ۲۰۰/۲۱۰۱۲ توسط سازمان امور مالیاتی کشور صادر گردید که به موجب آن حساب تعدیلات سنواتی (سود و زیان انباشته) طی دوره جاری می‌توان بدهکار یا بستانکار گردد، بنابراین مأمورین مالیاتی می‌بایست در زمان رسیدگی گردش حساب مذکور و مستندات مربوطه را اخذ و ضمن بررسی صحت محاسبات

مربوط، با رعایت مقررات و استثنائاتی آن را به درآمد مشمول مالیات عملکرد مورد رسیدگی اضافه نمایند. از آنجاکه تعدیلات سنواتی مستقیماً بر روی درآمد مشمول مالیات سال گذشته تأثیر می‌گذارد لذا در این پژوهش تعدیلات سنواتی به عنوان متغیری برای تقلب مالیاتی در نظر گرفته شده است.[۱۲]

در مورد معیارسوم؛ وجود تقلب با اهمیت، می‌تواند زمینه‌ساز اظهارنظر غیر مقبول توسط حسابرس صورتهای مالی باشد. اظهارنظر حسابرسی آنچنان با قوانین مالیاتی کشور ایران گره خورده است که به موجب تبصره ۱ ماده ۲۷۲ قانون مالیاتهای مستقیم، صورتهای مالی حسابرسی‌شده به شرح این ماده و مطالب مذکور در گزارش‌های حسابرسی و بازرسی قانونی مربوط که در چهارچوب مقررات این قانون تنظیم شده باشد، می‌تواند برای تشخیص درآمد مشمول مالیات اشخاص یادشده توسط ادارات مالیاتی مورد استفاده و استناد قرار گیرد.

نوع اظهارنظر حسابرس نقش اساسی در ارزیابی صحت و رعایت کامل قوانین و مقررات شرکت ایفا میکند. اظهارنظر تعدیل شده (غیرمقبول) نشان دهنده عدم رعایت صحیح قوانین از سوی شرکت می‌باشد. از این رو، شرکتهایی که از حسابرسان خود اظهارنظر غیر مقبول دریافت کرده‌اند نسبت به شرکتهای دیگر که اظهار نظر مقبول دریافت کرده‌اند از اجتناب و فرار مالیاتی بیشتری برخوردارند. مدیران شرکت‌ها در صورتیکه اظهارنظر غیرمقبول از سوی حسابرسان خود دریافت کنند از سوی سهامداران تحت فشار قرار خواهند گرفت لذا سعی مینمایند در سنوات آتی به گونه ایی اقدام کنند که از طریق گزارش مقبول اثر منفی گزارش های قبلی را برطرف نمایند. همچنین اخذ یک گزارش مقبول از سوی مدیران مالی یک رزومه و سابقه خوب برای آنها تلقی میشود در نتیجه سعی می کنند به منظور دریاقت یک گزارش مقبول، قوانین و مقررات را به نحو احسن اجرا نمایند. دریافت گزارش مقبول همچنین ممکن است منجر به ارائه پاداش از سوی مجامع عمومی سهامداران به مدیر هم شود. بنابراین میتوان نتیجه گرفت در صورتیکه اظهارنظر غیر مقبول حسابرس صادر شود احتمالا مدیران قوانین و مقررات را به طور کامل رعایت نکرده اند که بخشی از آن مربوط به محاسبات مربوط به سود ابرازی و مشمول مالیات و متعاقب آن هزینه مالیات می‌باشد. در نتیجه می توان گفت که اظهارنظر غیرمقبول احتمالا با افزایش اجتناب از مالیات همراه است. (همتی و همکاران، ۱۳۹۸)

بدین ترتیب بوسیله معیارهای فوق ابتدا لیستی از شرکت های پذیرفته شده در بورس اوراق بهادار تهران که بین سال های ۱۳۹۴ تا ۱۳۹۶ مرتکب تقلب در صورتهای مالی شده اند تهیه می‌شود و با توجه به در دسترس بودن اطلاعات شرکت ها تعداد شرکت های متقلب تعیین می شود. سپس تعداد شرکت های سالم را در همین دامنه زمانی تعیین و سپس بر اساس روش نمونه گیری تصادفی ساده نمونه های کافی انتخاب می شوند. با اعمال محدودیت های فوق در نهایت تعداد ۱۵ شرکت متقلب شناسایی و بدلیل متناظر بودن تعداد شرکتهای متقلب و غیر متقلب، به ازای هر شرکت متقلب، ۳ شرکت غیرمتقلب یعنی مجموعا ۴۵ شرکت در همان صنعت بعنوان نمونه آماری شرکتهای غیر متقلب انتخاب گردید.

متغیرهای مستقل در این پژوهش، نسبتهای مالی هستند؛ بطوری‌که با مطالعه و بررسی پژوهش‌های صورت گرفته در این زمینه از قبیل راسا کاناپیکین و گراندی (۲۰۱۵)، سانگ و همکاران (۲۰۱۴) و زارع بهنمیری (۱۳۹۵)، نسبتهای مالی مورد نیاز در این پژوهش منجر به انتخاب نهایی تعداد ۵۴ متغیر مستقل شد؛

سپس این متغیرها ازصورت‌های مالی شرکت‌های متقلب و سالم در سال‌های مورد مطالعه استخراج نموده و به‌وسیله مدل رگرسیون خطی متغیرهایی که دارای همبستگی معنادار با صورتهای مالی متقلبانه می‌باشد انتخاب نموده و از آن‌ها به عنوان متغیرهای ورودی مدل‌های یادگیری ماشین و الگوریتم ژنتیک استفاده می‌شود. جدول زیر میانگین، انحراف استاندارد و آزمون ANOVA را برای نسبتهای شرکتهای متقلب و غیرمتقلب گزارش می کند. آزمونهای تک متغیره به

چندین متغیر اشاره دارند که ممکن است در کشف شرکتهای متقلب مفید باشند. از بین ۵۴ متغیر مورد آزمون تعداد ۲۳ متغیر که در سطوح ۱ الی ۵ درصد معنادار هستند و سایر متغیرها فاقد معناداری مناسب بوده اند.

مدل پیشنهادی دارای مراحل مختلفی می باشد که به ترتیب عبارتند از:

۱) پیش پردازش داده ها؛
۲) انتقال داده ها؛
۳) انتخاب ویژگی های موثر با استفاده از الگوریتم ژنتیک و شبکه های عصبی؛
۴) آموزش و محاسبه اوزان الگوریتم درخت تصمیم؛
۵) ساخت درخت تصمیم گیری؛
۶) تبدیل درخت تصمیم گیری و بهینه سازی آن؛

که در ادامه به تشریح هر مرحله پرداخته می‌شود:

۸-۱- پیش پردازش داده ها

پس از استخراج نسبتهای مالی، به آماده سازی و پیش پردازش آن‌ها پرداخته می شود. در آماده سازی و پیش پردازش داده‌ها از روش‌های مختلفی استفاده می شود. اول این که برخی ویژگی‌ها دارای مقادیر منحصر به فرد هستند. این ویژگی ها نمی توانند دانش مفیدی را در مجموعه داده ایجاد کنند، لذا این مجموعه ویژگی ها باید از داده ها حذف شوند. به طور نمونه می توان به ویژگی نام شرکت، نام بورسی شرکت و صنعت آن اشاره نمود. همچنین ممکن است برخی تراکنش ها دارای مقادیر مفقود فراوان باشند. لذا این تراکنش ها نیز باید از مجموعه داده ها حذف شوند. از طرفی ممکن است، مقادیر برخی ویژگی ها دارای مقادیر پرت باشند لذا این مقادیر نیز باید در مجموعه داده اصلاح شوند. سپس از ابزار کشف آنومالی به منظور شناسایی و حذف داده هایی که در نقاط خارج از قانون مجموعه داده قرار دارند استفاده شده است. برای انجام برخی عملیات بر روی داده ها به عنوان ورودی، می بایست ویژگی‌هایی را از آن ها استخراج نمود. به طور معمول پیش از انتخاب و استخراج ویژگی‌ها، برخی عملیات پیش پردازش بر روی داده ها انجام می‌شود.

۸-۲- انتقال داده ها

در این قسمت داده ها در دامنه های درست قرار می‌گیرند. بدین معنا که داده ها باید به محدوده‌هایی که در سیستم مشخص شده است منتقل شوند و داده های خارج از محدوده، داده های مشکل‌دار بوده و می‌بایست حذف شوند. داده‌های می‌بایست در محدوده درست و منطقی قرار بگیرند. بدین معنا که برای مثال اگر فیلد سن وجود داشته باشد فردی که محدوده سنی بین ۵۵ تا ۷۰ دارد می بایست در سیستم بصورت خیلی پیر شود که این قسمت بصورت اتوماتیک از روی مجموعه داده ها تکمیل می شود.

۸-۳- انتخاب نسبت‌های مالی مرتبط با تقلب مالیاتی به وسیله الگوریتم ژنتیک و شبکه های عصبی

الگوریتم ژنتیک، با قابلیت قابل توجه در استنتاج معانی از داده‌های پیچیده، میتواند برای استخراج الگوها و شناسایی روش‌هایی که آگاهی از آنها برای انسان و دیگر تکنیک‌های کامپیوتری بسیار پیچیده و دشوار است به کار گرفته شود. الگوریتم ژنتیک به عنوان یکی از ابزارهای داده کاوی می تواند برای طبقه‌بندی مورد استفاده قرار گیرد. در این پژوهش از الگوریتم شبکه‌های عصبی پرسپترون چندلایه که به وسیله الگوریتم ژنتیک بهینه‌سازی شده، در جهت شناسایی ویژگی‌های موثرتر استفاده گردید تا در نتیجه نسبتهای مالی موثرتر بر تقلب شرکتها را شناسایی گردد. در این بخش از راهکاری که توسط لدسما و همکارانش (۲۰۰۸) ارائه شده استفاده می-گردد.

تعیین ساختار شبکه از گام‌های تاثیرگذار بر روی نحوه‌ی آموزش شبکه می‌باشد البته تعداد نورون بالا در شبکه، پیچیدگی آن را افزایش می دهد و ممکن است شبکه دچار بیش برازش شود که قابلیت پیش

بینی شبکه را تحت تاثیر قرار خواهد داد. به این جهت برای هر کدام از الگوریتم های آموزشی تعداد نورون ها از یک تا ۱۰ افزایش یافتند و داده ها در رنج ۱- تا ۱ نرمالیزه و از تابع انتقالی تانژانتی- سیگمویدی (tansig) استفاده گردید. همچنین از دو شاخص ارزیابی عملکرد شبکه خطای میانگین مربعات (MSE) و ضریب تبیین (R2) بهره گرفته شد. از رایج‌ترین الگوریتم های بهینه سازی که بر اساس دگرگونی بیولوژیکی بوجود آمده است، الگوریتم ژنتیک می‌باشد. جمعیت در الگوریتم ژنتیک شامل پاسخ های ممکن در فرم آرایه ای از کروموزوم ها باشد. وزن های شبکه توسط الگوریتم ژنتیک بهینه‌سازی شده و هر یک از جمعیتها به صورت تصادفی به عنوان وزن شبکه معرفی شدند. تابع MSE به عنوان تابع هزینه معرفی شد و کروموزوم های جمعیت سپس برای رسیدن به کمترین تابع هزینه مرتب می شوند. تعداد مشخصی از اعضای بهتر بر اساس کمترین هزینه به نسل بعدی منتقل می شوند. در این مرحله سه اپراتور الگوریتم ژنتیک (انتخاب، تقاطع و دگرگونی) برای تولید جمعیت نسل بعدی فعال می شوند. این سیکل تا رسیدن به جواب مطلوب ادامه یافته تا وزن های مطلوب شبکه حاصل شوند.

در نهایت در این قسمت تعداد ۲۳ نسبت مالی مورد بررسی و در صورت امکان تعداد آنها کاهش پیدا می کند که در این راستا سربار محاسباتی و همچنین نویز موجود نیز تا حد امکان کاهش پیدا می‌کند. به منظور سهولت در انجام عملیات، هر نسبت با استفاده از Att و یک عدد مشخص شده که در جدول زیر نشان داده شده است.

Att1	:	مجموع بدهیها / مجموع داراییها
Att2	:	نسبت جاری = دارایی های جاری / بدهی های جاری
Att3	:	نسبت آنی = دارایی های جاری – (موجودی کالا+پیش پرداخت) / بدهی های جاری
Att4	:	لگاریتم طبیعی (بهای تمام شده کالای فروش رفته)
Att5	:	سود خالص/ مجموع دارایی ها
Att6	:	سود خالص / فروش
Att7	:	فروش / مجموع دارایی ها
Att8	:	سود عملیاتی/ فروش
Att9	:	سود قبل از بهره و مالیات/ فروش
Att10	:	سود ناخالص / کل داراییها
Att11	:	سود قبل از بهره و مالیات/ کل داراییها
Att12	:	سود قبل از بهره و مالیات / بدهی های جاری
Att13	:	بدهی جاری/ (دارایی جاری- موجودی کالا)
Att14	:	موجودی کالا / بدهی جاری
Att15	:	وجه نقد / جمع بدهی ها
Att16	:	بدهی های جاری / جمع دارایی ها
Att17	:	سرمایه / جمع داراییها
Att18	:	موجودی کالا / فروش
Att19	:	حسابهای دریافتنی / فروش
Att20	:	فروش / دارایی ثابت
Att21	:	بهای تمام شده کالای فروش رفته / فروش
Att22	:	هزینه های عملیاتی / فروش
Att23	:	موجودی کالا / دارایی جاری

شکل (۴): لیست و عنوان نسبتهای استفاده شده

در روش پیشنهادی در ابتدا، کار انتخاب ویژگی صورت می‌گیرد و ویژگی‌های مؤثرتر در روش پیشنهادی انتخاب می‌شوند که برای مجموعه داده‌های ورودی می‌توان در شکل(۵) تاثیر ویژگی‌ها را مشاهده نمود که مرتب سازی شده است.

Att1 : 0
Att16 : 0
Att15 : 0
Att14 : 0
Att13 : 0
Att23 : 0
Att19 : 0
Att20 : 0
Att18 : 0
Att21 : 0
Att7 : 0
Att22 : 0
Att4 : 0
Att3 : 0
Att2 : 0
Att17 : 0
Att6: 0.0968621106031251
Att9: 0.097428655876255
Att12: 0.10260370571401
Att5: 0.106935401391966
Att11: 0.10726139663299
Att8: 0.115751269584137
Att10: 0.12294193114424

شکل (۵): مقدار وزن های آنتروپی اختصاص داده شده به نسبت‌های مالی

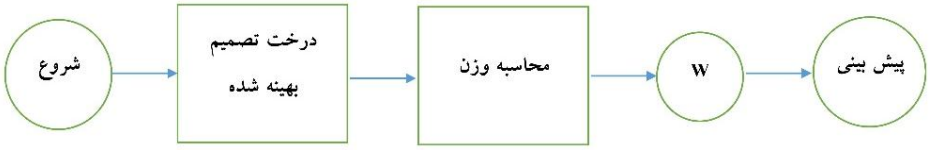
با توجه جدول فوق می توان نتیجه گرفت که نسبتهای مالی با ضریب صفر، تاثیری در خروجی ندارد و این در حالی است که نسبتهای مالی همچون سود قبل از بهره و مالیات به فروش، سود خالص به فروش، سود قبل از بهره و مالیات به بدهی های جاری، سود قبل از بهره و مالیات به کل داراییها، سود خالص به مجموع دارایی ها، سود عملیاتی به فروش و سود ناخالص به کل دارایی‌ها بیشترین تاثیر را در متقلب بودن و یا نبودن شرکت دارد. در اینجا می توان نسبتهای مالی بی تاثیر را حذف نمود تا الگوریتم قادر باشد با سرعت بیشتری داده کاوی را انجام دهد و همینطور در همین قسمت می‌توان دریافت که چه نسبتهایی بیشترین تاثیر را در متقلب بودن و یا نبودن شرکت دارد و چه عواملی باعث سوق دادن شرکت به سمت تقلب می باشد. یعنی می‌توان پارامترهای تاثیرگذار را در اینجا شناسایی نمود و روی این پارامترها توجه بیشتری صورت گیرد تا اینکه بتوان شرکت های متقلب را بهتر شناسایی نمود و تصمیمات مهمی را نیز در این راستا اتخاذ نمود.

۸-۴- آموزش و محاسبه اوزان الگوریتم های درخت تصمیم

در ابتدای کار درصدی از نسبتهای مالی مورد استفاده، برای محاسبه آموزش و محاسبه وزن مورد استفاده قرار می‌گیرد. در این قسمت از الگوریتم درخت تصمیم بهینه شده استفاده می‌شود. می‌توان نمایی از این مرحله را در شکل (۶) مشاهده نمود.

همانطور که در شکل (۶) مشاهده می شود الگوریتم پیشنهادی در ابتدا با استفاده از یک مجموعه داده به محاسبه وزن می‌پردازد. بدین ترتیب که الگوریتم با استفاده از ۷۰ درصد مجموعه داده‌های موجود آموزش می بیند و با استفاده از ۳۰ درصد باقیمانده مورد آزمون قرار می‌گیرد در نهایت با توجه به تعداد جواب‌های صحیح امتیاز و یا وزنی به آن تعلق می‌گیرد تا اینکه با استفاده از وزن متعلق، بتوان در مرحله بعدی وزنی را برای خروجی آن در نظر گرفت.

همانطور که در معماری نیز می‌توان مشاهده نمود بعد از محاسبه وزن که بصورت تقسیم تعداد جواب های درست به تعداد کل جواب‌های حدس زده شده می باشد، می‌توان میزان تاثیر گذاری الگوریتم را در خروجی نهایی بهتر تشخیص داد. در این روش بعد از محاسبه وزن‌ها، به ازای هر رکورد ، پیش بینی توسط درخت تصمیم بهینه صورت می‌گیردکه مقدار پیش-بینی شده می‌بایست در وزن آن الگوریتم ضرب شود و خروجی نهایی پیش بینی الگوریتم برابر است با جمع نتایج الگوریتم ضرب در وزن آن الگوریتم که بدین صورت نتیجه نهایی بدست می‌آید و دسته بندی درست صورت می‌گیرد.



شکل (۶): معماری استفاده شده در محاسبه وزن

۸-۵- ساخت درخت تصمیم‌گیری

درخت تصمیم گیری، درختی است که هر شاخه از آن به عنوان یک انتخاب می باشد. بدین ترتیب برای رفتن از گره ریشه به گره پایین‌تر می توان از شاخه هایی که به آن گره متصل هستند یکی انتخاب شود. در انتها هر یک از گره های انتهایی یا اصطلاحاً گره برگ تصمیمی را بازگو می‌کند. هر کدام از شاخه ها تا رسیدن به برگ دارای سناریویی می باشد که موجب تصمیمی می‌شود. در مدل پیشنهادی ارائه شده، از درخت تصمیم ID3 بهبود یافته استفاده شده است که بهبود انجام شده موجب سرعت عمل بالای آن شده است. درخت ID3 یک درخت تصمیم گیری می باشد که دارای یادگیری نیز می باشد و اولین بار توسط راس کوینلن[۱۳] مطرح شد. ایده الگوریتم ID3، ساخت درخت تصمیم گیری بالا به پایین می باشد که انتخاب گره در آن به وسیله جستجوی حریصانه از میان مجموعه ای از صفت ها می باشد. در این پژوهش برای اینکه قادر باشیم تا مفیدترین نسبت را از میان نسبت‌های مالی بیابیم که در کلاسه بندی مفیدتر باشد از الگویی بخصوص استفاده نمودیم. برای اینکه بتوان کلاسه‌بندی مفیدی را برای مجموعه یادگیری انجام داد، می بایست تعداد سوالات را کاهش داد یا می‌توان گفت می بایست عمق درخت تصمیم-گیری را کاهش داد. از این‌رو در این قسمت نیاز به تابعی است که قادر باشد تا متعادل‌ترین تقسیم را انجام دهد که در این صورت عمق درخت بسیار کاهش می یابد و گره ها بصورت متعادل در درخت تقسیم می‌شوند.

جدولی را در نظر بگیرید که دارای صفات و کلاسی از صفات نیز می‌باشد. در صورتی به این جدول همگن[۱۴] گفته می‌شود که تنها شامل یک کلاس باشد. اگر یک جدول دارای چندین کلاس باشد، در این حالت به آن ناهمگن[۱۵] گویند. توابع زیادی همچون آنتروپی، gini index و classification error برای سنجش میزان همگن‌پذیری وجود دارند. در این میان در اینجا از آنتروپی[۱۶] استفاده شده است.

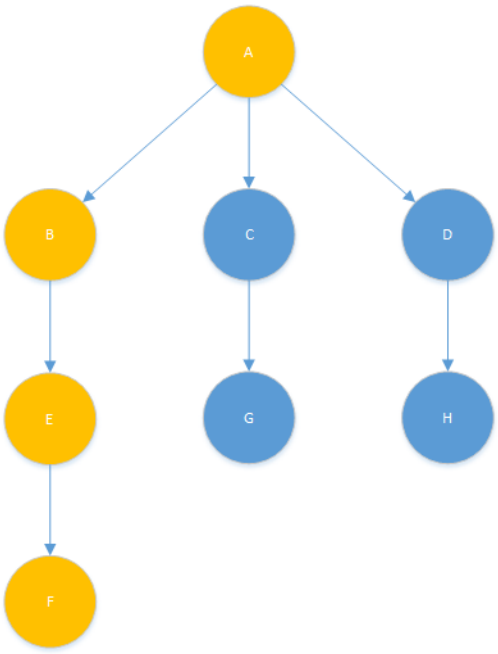
$$\text{Entropy} = \sum_{j} - p_j \log_2 p_j \qquad (۱)$$

آنتروپی یک جدول صفر است زیرا احتمال آن مقداری برابر یک است (تنها دارای یک کلاس باشد). آنتروپی زمانی به بیشترین مقدار خود می‌رسد که تمامی کلاس های موجود در جدول دارای احتمالی برابر باشند. آنتروپی را می توان به نوعی معیاری برای سنجش بی‌نظمی در نظر گرفت هر چه مجموعه منظم تر باشد و دارای گوناگونی کمتری باشد آنگاه آنتروپی آن کمتر است و به نوعی بی نظمی آن نیز کمتر است و برعکس. البته در اینجا چون ما در مرحله قبل یک کلاسه بندی ابتدایی را انجام دادیم تقریبا بی نظمی نیز پایین می‌باشد و این خود باعث سرعت عمل بالاتر مدل پیشنهادی در این پژوهش می شود زیرا این قضیه باعث می شود که عمق درخت تصمیم گیری کم شود و هر چه عمق این درخت کمتر باشد، سرعت تصمیم گیری نیز بیشتر می شود.

در این پژوهش، از آنتروپی استفاده گردید تا مقدار بی نظمی را برای هر صفت بدست آید. برای اینکه صفتی را در درخت تصمیم گیری انتخاب نماییم که در رتبه بالاتری از بقیه صفت ها باشد و به نوعی دارای اهمیت بالاتری از بقیه صفت ها باشد از فرمول (۱) استفاده گردید. با توجه به این فرمول ما آنتروپی همه صفات را در مجموعه S محاسبه نموده و مقدار صفت مجموعه A را از آن کسر می‌نماییم. مجموعه A، مجموعه همه صفات انتخاب شده از پدر تا به اینجا در یک مسیر خاص می باشد.

$$G(\mathrm{S},\mathrm{A}) = Entropy(S) - \sum_{v \in Values(A)} Entropy(\mathrm{v}) \qquad (۲)$$

برای درک بهتر فرمول (۲) می توان شکل زیر را مشاهده نمود.



شکل (۷): مثالی برای انتخاب صفت ها

همانطور که در شکل (۷) قابل مشاهده است، می بایست آنتروپی تمامی صفات را از آنتروپی صفات انتخاب شده تا به اینجای مسیر کاسته شود (یعنی باید G(S,F)=E(S)-(E(A)+E(B)+E(E)+E(F)) را بدست آورد). البته باید توجه شود که می بایست در مجموعه A صفاتی که تا به اینجای کار استفاده شده است به علاوه صفتی که می خواهیم قرار دهیم را محاسبه نماییم. بعد از اینکار از بین این مجموعه صفات باقیمانده که برای هر کدام فرمول (۲) را محاسبه نمودیم، صفتی را که دارای G بیشتری است را انتخاب نماییم. در این حالت اگر دو صفت دارای G برابر بودند که احتمال این پیشامد نیز کم نیست، می بایست به گره دو یا هر تعداد صفت که دارای بیشترین مقدار G هستند و باهم برابر نیز می باشند به گره مربوطه افزوده گردد؛ یعنی اگر برای مثال در گره ای دو صفت دارای G برابر بودند، آنگاه این دو به گره مربوطه دو فرزند می‌افزاییم و هر کدام از این صفت ها به عنوان یک فرزند این گره در نظر گرفته می‌شوند و بعد از این کار روند الگوریتم را برای هر یک از این گره ها ادامه می‌دهیم.

این کار باعث می شود تا صفت هایی که دارای آنتروپی بیشتری هستند را بیابیم چراکه این صفت ها تاثیر بیشتری را در تصمیم نهایی می گذارند. این روند جلو رفتن در درخت تصمیم گیری تا جایی ادامه می یابد که در هر مسیر دیگر صفتی باقی نمانده باشد. در این حالت درخت تصمیم گیری کاملا ساخته شده است و به پایان رسیده است.

۹- ارزیابی مدل پیشنهادی

در این بخش مدل پیشنهادی مورد بررسی قرار گرفته و با الگوریتمی معروف به نام الگوریتم آدابوست مورد مقایسه قرار می‌گیرد. همانطور که در قسمت قبل بیان شد مدل ارائه شده مبتنی بر ID3 است که با شبکه عصبی پرسپترون چند لایه بهینه سازی شده با الگوریتم ژنتیک ترکیب گردید و داده‌های مورد استفاده در آن، نسبتهای مالی مربوط به صورتهای مالی سه سال از شرکت‌های متقلب و غیر متقلب می باشد.

داده ها در ابتدا توسط برنامه به فرمت مناسب برای تحلیل قرار می گیرد یا به عبارتی پیش پردازش ابتدایی صورت می گیرد فایلی با فرمت ARFF ایجاد می باشد که ساختاری مناسب و استاندارد برای تحلیل می باشد.

پیاده سازی در برنامه Visual Studio 2017 و با زبان برنامه نویسیC# صورت گرفته است و در حین کار از کتابخانه‌هایی مانند weka و zedgraphکمک گرفته شده است. سیستم مورد استفاده در اینجا دارای سیستم عامل Windows 10 ، دارای ۶ گیگRAM و Corei7 می باشد. در ادامه می توان نتایج پیاده سازی را برای این الگوریتم ها مشاهده نمود.

```
--------AdaBoost----------
confusionMatrix:
[0,0] = 8 [0,1]=7
[1,0] = 38 [1,1]=111
Correct Prediction Percent = 66.2999999999999%
InCorrect  Prediction Percent = 33.7000000000001%
MeanAbsoluteError(MAE) = 0.3564377854847
MeanSquaredError(MSE) = 0.4676490987127
RelativeAbsoluteError(REA) = 88.72113460925
Correct Prediction Number = 119
InCorrect Prediction Number = 61
TP: 111
FP: 17
FN: 44
TN: 8
```

جدول (۱): خروجی مربوط به الگوریتم آدابوست

```
--------MyAlgorithm----------
confusionMatrix:
[0,0] = 11 [0,1]=7
[1,0] = 26 [1,1]=136
Correct Prediction Percent = 81.39999999999%
InCorrect Prediction Percent = 18.60000000001%
MeanAbsoluteError(MAE) = 0.223245665789764
MeanSquaredError(MSE) = 0.35983245468912
RelativeAbsoluteError(REA) = 60.621457548094
Correct Prediction Number = 147
InCorrect Prediction Number = 33
TP: 136
FP: 7
FN: 26
TN: 11
```

جدول (۲): خروجی مربوط به الگوریتم پیشنهادی

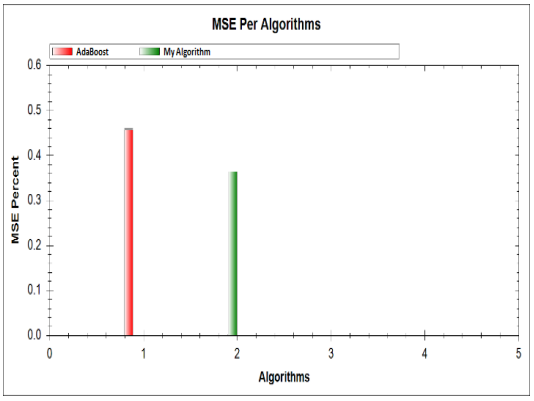
با توجه به جداول خروجی فوق می‌توان به وضوح مشاهده نمود که الگوریتم پیشنهادی با ۸۱,۴٪ دقت دارای بالاترین دقت صحت و با ۱۸,۶٪ اشتباه دارای کمترین میزان اشتباه می‌باشد. در جدول شماره (۳) به مقایسه مدل ارائه شده و الگوریتم آدابوست پرداخته می‌شود.

جدول (۳): مقایسه درصد صحت و خطای پیش بینی ها

نام الگوریتم	درصد صحت پیش بینی‌ها	درصد خطای پیش بینی‌ها
الگوریتم پیشنهادی	٪۸۱,۴	٪۱۸,۶
الگوریتم آدابوست	٪۶۶,۳	٪۳۳,۷

با توجه به دقت‌های بدست آمده، کاملا قابل مشاهده می باشد که مدل پیشنهادی، از الگوریتم آدابوست بهتر عمل کرده است. در اینجا می توان کاملا مشاهده کرد که در روش ترکیبی پیشنهادی بیان شده می توان با بهبود ایجاد شده درID3 که روند بهبود آن در قسمت قبل بیان شد و ترکیب آن با شبکه عصبی پرسپترون چند لایه به دقت بالاتر و خطای پایین‌تری در شناسایی تقلب مالیاتی دست یافت. این درخت دارای کمترین ارتفاع ممکن می باشد و از اینرو دارای سربار پایین‌تری می باشد.

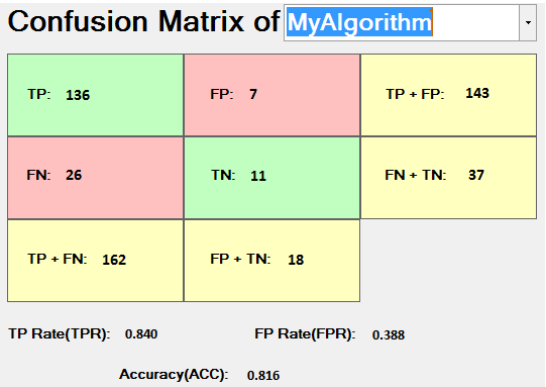
همچنین از آنجا که در علم داده کاوی، معیار خطای میانگین مربعات (MSE)[۱۷] دارای اهمیت بسیار بالایی می باشد و از اعتبار بسیار زیادی برخوردار است به بررسی آن در شکل زیر می‌پردازیم. این معیار میانگین خطای مربعات را بدست می آورد و با استفاده از این معیار می توان به طور دقیق پیش بینی را سنجید و میزان خطای پیش بینی را محاسبه نمود.



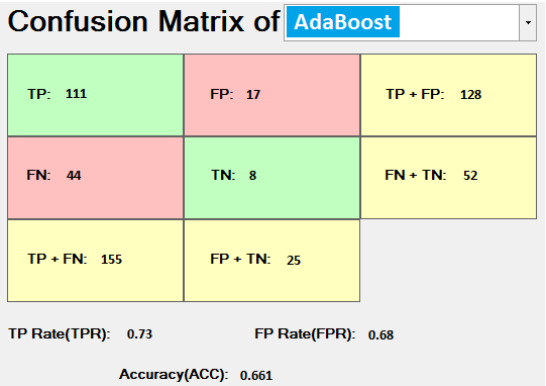
شکل شماره (۸): میزان معیار MSE در میان الگوریتم پیشنهادی و الگوریتم آدابوست

همانطور که ملاحظه شد، الگوریتم پیشنهادی دارای MSE پایین‌تری نسبت به الگوریتم آدابوست می‌باشد که نشان دهنده این است که میزان خطای پیش‌بینی در مدل ارائه شده، کمتر از الگوریتم آدابوست می‌باشد.

در ادامه جداول خروجی ماتریس درهم ریختگی[18] مربوط به مدل پیشنهادی و الگوریتم آدابوست در این تحقیق نشان داده شده است.



شکل شماره (۹): ماتریس درهم ریختگی الگوریتم پیشنهادی



شکل شماره (۱۰): ماتریس درهم ریختگی الگوریتم آدابوست

همانطور که می‌توان مشاهده نمود که در روش پیشنهادی مقادیر TP و TN بیشتر و در کنار آن نیز دارایFP و FN کمتری از الگوریتم آدابوست می باشد. زیرا هر چه یک الگوریتم دارایTP و TN بیشتری باشد یعنی متقلب بودن و یا نبودن های در مجموعه داده‌های تست را به مقدار بیشتری درست تشخیص داده است و FP و FN نشان دهنده برعکس این قضیه یعنی پیش‌بینی اشتباه می باشد.

۱۰- بحث و نتیجه گیری

امروزه دانش به عنوان یک منبع ارزشمند و استراتژیک و نیز یک دارایی برای ارزیابی و پیش بینی مطرح است و ارائه این راهکارها در زمینه کشف شرکت های متقلب باعث افزایش دقت و همچنین کاهش نیروی کار موثر و همیشگی برای بررسی و تشخیص شرکت های دارای تقلب مالیاتی می شود یعنی می توان از مدلی مانند مدل پیشنهادی بصورت تمام وقت به بررسی و کشف شرکت های دارای تقلب مالیاتی پرداخت و این نیازمند نیروی کار انسانی نمی-باشد بلکه خود سیستم می تواند بصورت هوشمندانه تشخیص را انجام داده و اطلاع رسانی را انجام دهد.

بزرگترین طرح تاریخ مالیاتی کشور به نام طرح جامع مالیاتی بر مبنای شناسایی ریسک تقلب مالیاتی مودیان با استفاده از نسبتهای مالی پایه گذاری شده است که در این پژوهش با مطالعه بر روی نسبتهای مالی مرتبط با تقلب مالیاتی راهکاری برای ارزیابی و پیش بینی تقلب‌های مالیاتی شرکت‌ها ارائه شد و مشاهده شد که روش ارائه شده دارای عملکرد مناسبی بود که روش پیشنهادی نسبت به الگوریتم آدابوست بهبود عملکرد داشته است. داده های استفاده شده در این تحقیق مربوط به ۶۰ شرکت در طی ۳ سال می‌باشد بدین ترتیب، داده های مورد بررسی در این تحقیق دارای ۱۸۰ رکورد بود. داده ها در ابتدا مورد پیش پردازش قرار گرفته و انتقالاتی نیز روی داده‌ها صورت گرفت تا اینکه داده‌ها به داده‌های ورودی مورد نیاز الگوریتم پیشنهادی تبدیل شوند. نتایج بدست آمده کاملاً بهبود روش پیشنهادی را نشان می دهد. لذا با توجه به نتایج دریافتی مدل ارائه شده با ۸۱,۴٪ دقت دارای دقت و صحت بالاتر و با ۱۸,۶٪ اشتباه دارای میزان اشتباه کمتری نسبت به الگوریتم آدابوست در شناسایی شرکت‌های متقلب مالیاتی می‌باشد که می‌توان از آن جهت پیش‌بینی و یا به عنوان نماینده تقلب مالیاتی در بررسی‌ها و تحقیقات مختلف پیشنهاد نمود.

نتایج این پژوهش در رابطه با افزایش دقت در کشف تقلبات با استفاده از روش های داده کاوی و شبکه عصبی با پژوهش های دیموید و دیگران (۲۰۱۸)، لاری دشتبیاز و دیگران (۲۰۱۵)، ووا، اوو، لین، چنگ و ین (۲۰۱۳)، راویسانکار، راوی، رائو و

بس (۲۰۱۱)، ژو و کاپور (۲۰۱۱)، کِرکوس، اسپاتیس و مانولوپولوس (۲۰۰۷)، موسایی و دیگران (۱۳۹۸)، نمازی و صادق زاده مهارلوئی (۱۳۹۷)، تقوی فرد و همکاران (۱۳۹۶)، منطبق می باشد.

به پژوهشگران آتی پیشنهاد میشود که با توجه به اینکه در این پژوهش از آنتروپی استفاده شده است از روش های دیگر نیز برای برازش مدل استفاده نمایند و یا این روش را با روش های دیگری ادغام نمایند. برای مثال در صورتی که روش این پژوهش را با روشی مانند Gain که تابع ارزش می باشد ادغام نماییم، به احتمال زیاد دارای عملکرد بهتری می باشد چراکه آنتروپی نیز دارای معایبی می باشد ولی سرعت آن بالاست و در اینجا هم ما به دنبال روشی بودیم که دارای سرعت بالا باشد ولی می توان با ادغام این روش و یا روش های جایگزین دقت این روش ارائه شده را به شدت افزایش داد. همچنین می توان روش پیشنهادی را با بهبود در الگوریتم ۴,۵C نیز استفاده نمود (بدین ترتیب که، روش پیشنهادی را روی ۴,۵C با بهبودی مشابه بهبودی که روی ID3 در این تحقیق انجام شد)، تا عملکرد آن مورد آزمون قرار گرفته و صحت آن بررسی شود.

فهرست منابع

* اعتمادی، حسین و حسن زلقی، (۱۳۹۲)، کاربرد رگرسیون لجستیک در شناسایی گزارشگری مالی متقلبانه، مجله دانش حسابرسی، سال ۱۳، شماره ۵۱، ۱۴۵-۱۶۳.
* پور حیدری، امید و مجتبی گل محمدی شورکی (۱۳۹۴)، تأثیر ریسک وضعیت مالیاتی شرکت بر حق‌الزحمه حسابرسی، بررسی های حسابداری و حسابرسی، شماره ۲۲، صص ۳۱۸-۳۰۱.
* تشدیدی الهه، سحر سپاسی، حسین اعتمادی و عادل آذر (۱۳۹۸)، ارائه رویکردی نوین در پیش-بینی و کشف تقلب صورتهای مالی با استفاده از الگوریتم زنبور عسل، مجله دانش حسابداری، دوره دهم، شماره ۳، پاییز ۱۳۹۸، ۱۳۹-۱۶۷
* تقوی فرد، سید محمدتقی، ایمان رئیسی وانانی و ریحانه پناهی (۱۳۹۶)، تحلیل آینده نگر تشخیص فرار مالیاتی مؤدیان مالیات بر ارزش افزوده با استفاده از الگوریتم‌های طبقه‌بندی و خوشه بندی، پژوهشنامه مالیات، شماره ۳۵، ۱۱-۳۵
* داغمه چی فیروزجایی، علی (۱۳۹۳)، پاسخگویی در گزارشگری مالی: کشف شرکتهای متقلب، پایان نامه کارشناسی ارشد، رشته حسابداری، دانشگاه مازندران.
* زارع بهنمیری، محمد جواد (۱۳۹۵)، پیش بینی تقلب در صورتهای مالی با استفاده از نسبت های مالی، پایان نامه دکتری، رشته حسابداری، دانشگاه مازندران.
* صدر موسوی، میرستار و اکبر رحیمی (۱۳۸۸)، مقایسه نتایج شبکه‌های عصبی پرسپترون چندلایه با رگرسیون چندگانه در پیش‌بینی غلظت ازن در شهر تبریز، پژوهش‌های جغرافیای طبیعی، شماره ۷۱، ۶۵-۷۲
* عرب صالحی، مهدی و مجید هاشمی (۱۳۹۴)، تأثیر اطمینان بیش از حد مدیریتی بر اجتناب مالیاتی، بررسی های حسابداری و حسابرسی، شماره ۲۲، ۱۰۴-۸۵
* عرب مازار، علی اکبر و علی دهقانی (۱۳۸۸)، برآورد کارایی مالیات بر درآمد مشاغل و اشخاص حقوقی در استان‌های کشور، مجله پژوهش‌های مالیاتی، فصلنامه پژوهشنامه مالیات، سال ۱۷، شماره ۷، صص ۶۴-۴۵
* فرقاندوست حقیقی، کامبیز و فرید بروازی (۱۳۸۸)، بررسی کاربرد روش های تحلیل در ارزیابی ریسک تحریف صورتهای مالی(تقلب مدیریت)، مجله دانش و پژوهش حسابداری، شماره ۱۶، ۱۸-۷۰.
* قانون مالیاتهای مستقیم، مصوب ۱۳۹۴/۴/۳۱.
* محمدموسایی جابر، بابک جمشیدی نوید، مهرداد قنبری و فرشید خیراللهی (۱۳۹۸)، تدوین مدل کشف تقلب با استفاده از رویکرد ترکیبی برپایه مدل تحلیل عاملی و روش شبکه عصبی مصنوعی

در شرکت های پذیرفته شده در بورس اوراق بهادار تهران، مجله حسابداری مدیریت، دوره ۱۲، شماره ۴۲، پاییز ۱۳۹۸، ۷۵-۸۷.

* مهام، کیهان و ابوالفضل ترابی (۱۳۹۱)، ارائه مدل رتبه بندی ریسک در تقلب گزارشگری مالی، همایش جهاد اقتصادی (با تاکید بر تولید ملی، حمایت از کار و سرمایه ایرانی)، دانشگاه مازندران.
* مهرانی، ساسان و سیدجلال سیدی، (۱۳۹۳)، بررسی تأثیر مالیات بر درآمد و حسابداری محافظه کارانه بر اجتناب مالیاتی شرکت‌ها، فصلنامه علمی پژوهشی دانش حسابداری و حسابرسی مدیریت، شماره دهم، ۳۳-۱۳.
* نمازی، محمد و محمدصادق زاده مهارلوئی (۱۳۹۶)، پیش بینی فرار مالیاتی با استفاده از الگوریتم های داده‌کاوی درخت تصمیم، فصلنامه حسابداری مالی، سال نهم، شماره ۳۶، زمستان ۹۶، ۷۶-۱۰۰
* همتی، هدی و محمدعلی جابری و کریم ایمانی و محمدعلی شامحمدی، (۱۳۹۸)، بررسی رابطه نوع اظهارنظر و اندازه موسسه حسابرسی با اجتناب از مالیاتی با استفاده از روش دو مرحله ای همکن، نشریه تحقیقات حسابداری و حسابرسی، پاییز ۱۳۹۸، شماره ۴۳، ۱۷۱-۱۸۶.
* همراز، سعید (۱۳۸۵)، ماشین یادگیری همه منظوره، پایان نامه کارشناسی، رشته مهندسی کامپیوتر- نرم افزار، دانشگاه علم و صنعت ایران
* A.V Senthil Kumar et. al. 2013, "Diagnosis of heart disease using Advanced Fuzzy Resolution Mechanism" International Journal of Science and Applied Information Technology (IJSAIT), Vol. 2, No.2, Pages: 22-30 (2013).
* Didimo, W., L. Giamminonni, G. Liotta, F. Montecchiani, and Pagliuca, D. (2018). A Visual Analytics System to support Tax Evasion Discovery. Decision Support Systems.
* Hanlon, M, Heitzman, Sh. (2010). "A review of tax research". Journal of Accounting and Economics 50. Pp. 127-178.
* Hao Wang, Chengzhi Mao, Hao He, Mingmin Zhao, Tommi S. Jaakkola, Dina Katabi, "Bidirectional Inference Networks: A Class of Deep Bayesian Networks for Health Profiling", Machine Learning (stat.ML); Artificial Intelligence (cs.AI); Computer Vision and Pattern Recognition (cs.CV); Machine Learning (cs. LG),2019.
* https://chistio.ir
* http://hkamal.persiangig.com/document/genetic.Doc
* http://farsithesis.ht3.ir/farsithesis/41484/html
* Kantesh Kumar Oad, Xu DeZhi & Pinial Khan Butt et. al, "A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease". Global Journal of Computer Science and Technology: C Software & Data Engineering, Volume14
* Kim, Yeonkook J., Baik, Bok. Cho, Sungzoon. (2016). Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning, Expert Systems with Applications, No. 62, pp. 32-43.
* Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent. *Expert Systems with Applications, 32*(4), 995–1003
* Lari. Dashtbayaz, Mahmoud. (2015). Data search and discovery process for financial statement fraud, Research Journal of Finance and Accounting, Vol.6, No.3.
* Sergio Ledesma, Gustavo Cerda, Gabriel Avina, Donato Hernandez and Miguel Torres, Feature Selection using Artificial Neural Networks, A. Gelbukh and E.F. Morales (Eds.): MICIAI 2008, LNAI 5317, pp. 351-359.
* Ojeme Blessing Onuwa et. all, "Fuzzy Expert System for Malaria Diagnosis"., An International Open Free Access, Peer Reviewed Research Journal, Published by: Oriental Scientific Publishing Co., India. June2014, Vol.7, No. (2): Pgs. 273-284 [ISSN: 0974-6471].
* Ravisankar, P., Ravi, V., Rao, R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data miningtechniques. Decision Support Systems,50(2), 491–500.
* Rasa Kanapickiene, Zivile Grundiene, (2015), The Model of Fraud Detection in Financial Statements by Means of Financial Ratios, 20th International Scientific Conference Economics and Management, 2015
* Sergio Ledesma, Gustavo Cerda, Gabriel Avina, Donato Hernandez, and Miguel Torres, "Feature Selection Using Artificial

یادداشت‌ها

[1] AdaBoost
2 Genetic algorithm
3 Artificial Neural Networks
4 Multi Layer Perceptron
5 Self Organizing Map
6 Learning Vector Quantization
7 Hopfield Network
8 Time Delay Neural Network
9 Radial Basis Functions
[10] Decision tree
[11] Information Theory
[12] دستورالعمل رسیدگی به حساب تعدیلات سنواتی، شماره ۲۰۰/۲۱۰۱۲ مورخ ۱۳۹۱/۱۰/۲۵، سازمان امور مالیاتی کشور
[13] Ross Quinlan
[14] homogenous
[15] heterogeneous
[16] Entropy
[17] Mean squared error
[18] *confusion matrix*

Neural Networks", A. Gelbukh and E.F. Morales (Eds.): MICAI 2008, LNAI 5317, pp. 351–359, 2008.

* Seyedhossein Naslmosavi, Arezoo Aghaei Chadegani and Mohammadghorban Mehri. (2011), Comparing the Ability of Bayesian networks and Adaboost for Predicting Financial Distress of Firms listed on Tehran Stock Rxchange (TSE), Australian Journal of Basic and Applied Sciences, 5(10): 629-634, 2011
* Shin, K & Lee, Y. (2002). A Genetic Algorithm Application in Bankruptcy Prediction Modeling. Expert Systems with Applications, Vol. 23, Issue. 3, pp.321-328.
* Sudan Chen. (2016). Detection of fraudulent financial statements using the hybrid data mining approach, Springer Plus, No. 5:89.
* Vanhoeyveld Jellis, David Martens, Bruno Peeters. (2020), Value-added tax fraud detection with scalable anomaly detection techniques. Applied Soft Computing Volume 86, January 2020.
* Watts, R. (2003a). Conservatism in Accounting Part I: Explanations and Implications.AccountingHorizons 17 (3): 207-222.
* Xin-Ping Song, Zhi-Hua Hu, Jian-Guo Du, Zhao-Han Sheng, 2014, Application of Machine Learning Methods to Risk Assessment of Financial Statement Fraud: Evidence from China, Journal of Forecasting, J. Forecast. 33, 611–626 (2014)
* Ziming Yin, Yinhong Zhao, Xudong Lu, and Huilong Duan, 2014, Screening of Alzheimer's Disease Based on Multiple Neuropsychological Rating Scales, Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine, Volume 2015.
* Zhou, W., & Kapoor, G. (2011). Detecting evolutionaryfinancialstatement fraud. Decision Support Systems, 50(3), 570-575.