

پیش‌بینی نوع اظهار نظر حساب‌رسان با استفاده از طبقه‌بندی کننده تجمیعی لاجیت‌بوست:

با تاکید بر انتخاب متغیرهای پیش‌بین بهینه

سمانه قوهستانی<sup>۱</sup>

مصطفی کاظم‌نژاد<sup>۲</sup>

### چکیده

در اغلب پژوهش‌های انجام شده در بورس اوراق بهادار تهران، از روش‌های انفرادی پیش‌بینی (خطی و در پژوهش‌های اندکی نیز از روش‌های هوش مصنوعی) برای پیش‌بینی نوع اظهار نظر حساب‌رسان استفاده شده است. با توجه به مزایای طبقه‌بندی کننده‌های تجمیعی نسبت به طبقه‌بندی کننده‌های انفرادی، پژوهش حاضر به پیش‌بینی نوع اظهار نظر حساب‌رسان شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از طبقه‌بندی کننده تجمیعی لاجیت‌بوست می‌پردازد و عملکرد این روش با طبقه‌بندی کننده‌های متداول رگرسیون لجستیک و شبکه‌های عصبی مصنوعی مقایسه می‌شود. به منظور ارزیابی عملکرد این طبقه‌بندی کننده‌ها، از معیارهای میانگین دقت و خطاهای نوع اول و دوم استفاده شده است. یافته‌های تجربی مربوط به بررسی ۱۲۱۴ سال-شرکت پذیرفته شده در بورس اوراق بهادار تهران در بازه زمانی ۱۳۸۶ الی ۱۳۹۳ حاکی از عملکرد بهتر این طبقه‌بندی کننده‌ها نسبت به رگرسیون لجستیک و شبکه‌های عصبی مصنوعی است.

**واژه‌های کلیدی:** پیش‌بینی نوع اظهار نظر حساب‌رسان، طبقه‌بندی کننده تجمیعی لاجیت‌بوست، شبکه‌های عصبی، رگرسیون لجستیک.

---

<sup>۱</sup> دانشجوی دکتری حسابداری دانشگاه شیراز، حسابرس دیوان محاسبات فارس  
<sup>۲</sup> دکتری حسابداری دانشگاه شیراز

## ۱. مقدمه

وجود اطلاعات مالی شفاف و قابل مقایسه، رکن رکن پاسخگویی و تصمیمات اقتصادی آگاهانه و از ملزومات تبیین و توسعه اقتصاد ی به شمار میرود. سرمایهگذار اینک تصمیمبخرید یا فروش سهام دارند، بانکها و موسسات مالی و اعتباری که در مورد تأیید تقاضای وام تصمیم میگیرند و دولتی که بر ایو صولدر آمد مالیاتیبهاظهار نامه های مالیاتی اتکاء می کند، همگی بر اطلاعاتی که می کنند که توسط دیگران تهیه شده است. در بیشتر موارد، منافع تهیه کنندگان اطلاعات با منافع استفاده کنندگان از آن اطلاعات متفاوت است (فرضیه تضاد منافع در تئوری نمایندگی). نیاز به وجود حسابرسان مستقل در چینیاستد لایمستتر است؛ یعنی افراد یا صلاحیت و صداقت حرفه ای که بتوانند مشخص کنند اطلاعات مالی که بر آن اتکاء میشود، تصویری مطلوب و کامل از واقعیت را مستیخیر؟ (ارباب سلیمانی و نفری، ۱۳۸۵: ۱). کمیته مفاهیم بنیادی حسابرسی جامعه حسابداران آمریکا (۱۹۷۳) نیز در خصوص نقش حسابرسی به ایجاد ارزش برای اطلاعات توسط حسابرس اشاره کرده است. حسابرس مستقل از طریق اعتبار دهی به صورت های مالی، تضمین قابلیت اتکاء و تأیید کیفیت اطلاعات مالی، موجب حمایت از حقوق تمام ذینفعان شرکت میشود (اشبو<sup>۳</sup> و وارفیلد<sup>۴</sup>، ۲۰۰۳: ۲-۳). محصول نهایی حسابرسی یک واحد تجاری، گزارشگری است که در آن، حسابرسان نسبت به صورت های مالی صاحبکار اظهار نظر میکنند. شهادت دهی حسابرسان در قالب اظهار نظر حرفه ای آنها بیان میشود (نیکخواه آزاد، ۱۳۷۹: ۱۱). از طرفی یافته های پژوهش های گذشته حاکی از آن است که گزارش حسابرس، نوع و محتوای آن دارای محتوای اطلاعاتی و پیامدهای اقتصادی با اهمیت است و می تواند بر قیمت یا بازده سهام (چنوزائو، ۲۰۰۰؛ شباهنگ و خاتمی ۱۳۷۷ و حساس یگانه و یعقوبی منش، ۱۳۸۲)، پیش بینی ورشکستگی (شورورزی و همکاران ۱۳۹۰)، پاداش مدیران (دی آنجلو، ۱۹۸۱) و تصمیم گیری اعتبار دهندگان در خصوص اعطای اعتبار (فیرس، ۱۹۸۰) مؤثر باشد. بنابراین، پیش بینی دقیق نوع و محتوای گزارش حسابرسی، اطلاعات سودمندی در تصمیم گیری استفاده کنندگان از صورت های مالی محسوب می شود. به رغم عملکرد بهتر طبقه بندی کننده های تجمیع نسبت به طبقه بندی کننده های انفرادی در سایر پژوهش های حسابداری و مالی، پژوهشی که به پیش بینی نوع اظهار نظر حسابرسان، با استفاده از این روش ها پرداخته باشد، مشاهده نشد. در این راستا، پژوهش حاضر به بررسی عملکرد این روش ها در پیش بینی نوع اظهار نظر حسابرسان در شرکت های پذیرفته شده در بورس اوراق بهادار تهران می پردازد.

## ۲. مبانی نظری پژوهش

### ۲.۱. گزارش حسابرس و اهمیت پیش بینی آن

اطلاعات مبنای اساسی تصمیمگیری است. جهت دار بودن، گمراهکنندگی، نامربوط بودن و ناقص بودن اطلاعات میتواند منجر به تصمیمگیری غلط شود و به تصمیمگیرنده زیان وارد کند. در شرایطی که این تصمیمگیرها پیامدهای مهم اقتصادی داشته باشند، تصمیمگیرنده اهمیت بیشتری برای کسب اطمینان از کیفیت اطلاعات دریافت شده قائل میشود (حساس یگانه و یعقوبی منش، ۱۳۸۲: ۳۰). اطلاعات در صورتی دارای پیامدهای مالی و اقتصادی هستند که مربوط و قابل اتکاء باشند (نوروش و همکاران، ۱۳۸۹: ۱۳۸۹).

<sup>۳</sup> - Ashbaugh

<sup>۴</sup> - Warfield

۴۲). از سوی دیگر استفاده کنندگان از صورتهای مالی، هنگامی میتوانند به اطلاعات مالی منعکس در صورتهای مالی اتکاء کنند که شخصی مستقل، ذیصلاح، مطلع و بیطرف نسبت به میزان اعتبار این اطلاعات، نظر حرفهای ارائه کرده باشد (مرادی و فخرآبادی، ۱۳۸۸: ۹۰). حسابرسبها دعاهای یکتوسط شخصدیگری در قالب صورتهای مالی تهیه شده است، اعتبار میدهد و بدینوسیله قابلیت اتکا یا اطلاعات به کار رفتهدر تصمیمهای اقتصادی را افزایش میدهد (نیکخواه آزاد، ۱۳۷۹: ۱). محصول نهایی حسابرسی یک واحد تجاری، گزارشگری است که در آن، حسابرسان نسبت به صورتهای مالی صاحبکار ارائه میکنند (نیکخواه آزاد، ۱۳۷۹: ۱۱). اجرائی حسابرسی، افزونبر افزایش اعتماد و اطمینان نسبت به اعداد و ارقام مندرج در صورت های مالی به عنوان معیار ارزو یا بعلی عملکرد نمایندگان، از طریق کش فاشتبا هات، اعتماد پذیر بصورت های مالی را نیز افزایش میدهد.

هر سرمایه گذار منطقیو عقلایی به طور معمولی تواند میزان اعتماد پذیر یا اطلاعات تهیه و ارائه شده از سوی واحد اقتصادی را با اطلاعات تقریباً نمقایسه کن د. صاحبان سهامو اعتبار دهندگان، با اتکابراطلاعات مالی منباعتصا دیخودر اختصاصی دهند.

تخصیص منباعتاز سوی این افراد و گروه هادر نهایتباعث ایجاد منافعبرای آندستهاز واحد های اقتصادی می شود که هاطلاعات مالی قابلا اعتماد تر برارائه کرده ان د (امیراصلانی، ۱۳۸۴: ۷۳).

با توجه به مطالب پیش گفته، نقش حسابرسی، افزایش اعتبار و منصفانه بود ناطلاعات ارائه شده در صورتهای مالی است

حسابرسی با ایفای نقش خود می تواند بر سود مند یو قابلیت اتکا یا اطلاعات به عنوان بخش سازو ویژگی های کیفی آن، تاثیر گذارد

به دلیل توانایی حسابرسیدر افزایش اعتبار و منصفانه بود ناطلاعات، انتشار گزارش های حسابرسی می تواند منجر بهمخبرهاخبار واطلاعات دربار هشرک تبه بازار سرمایه بشود و در نتیجه واکنش بازار سرمایه را در پیداشتها شد . واکنش

بازار سرمایه بهمحتوایاطلاعاتی گزارش های حسابرسی، موجب تغییر در ارزش بازار شرکت هادر بازار سرمایه می شود

بنابراین، حسابرسی با ارائه گزارش های سود مند و انتشار آن هادر بازار سرمایه می تواند بر تصمیم گیری مشارکنندگان در بازار سرمایه تاثیر گذارد

وزمینهر ابرای اتخاذ تصمیم گیری های صحیح اقتصاد یفر ا همکنند (بنی مهد و علی احمدی، ۱۳۹۲: ۱۴). یافته های برخی از پژوهش های تجربی نیز حاکی از اهمیت گزارش حسابرسی در تصمیم گیری و پیامدهای اقتصادی نوع اظهار نظر حسابرس است که در بخش مقدمه به برخی از آن ها اشاره شد. بنابراین، با توجه به محتوای اطلاعاتی گزارش حسابرسی، پیش بینی نوع اظهار نظر حسابرسان حائز اهمیت است.

## ۲.۲. انتخاب متغیرهای پیش بین بهینه و اهمیت آن در پیش بینی

از دیدگاه نظری، یادگیری از تعداد متغیرهای پیشین بیشتر باعث میشود تا دقت پیش بینی بالا رود. با این وجود، شواهد تجربی بیانگر آن است که این امر همواره صادق نیست؛ زیرا تمام متغیرها، برای تشخیص و پیش بینی طبقه دادهها مهم نیستند و یا برخی از آنها اساساً با طبقه دادهها بیارتباط هستند (لیندنباوم و همکاران، ۲۰۰۴). با توجه به اینکه عاملهای بسیاری از جمله کیفیت دادهها در موفقیت یک الگوریتم یادگیری موثر است، اگر دادهها حاوی متغیرها و یا اطلاعات تکراری و نامربوط باشند و یا حاوی اطلاعات نویزی و نامطمئن باشند، اخذ دانش از آن دادهها مشکل می گردد (هال، ۲۰۰۰). افزون بر این، کاهش تعداد متغیرهای پیش بین نامربوط یا اضافی<sup>۱۱</sup> (تکراری)، علاوه بر کاهش زمان اجرای الگوریتمهای آموزشی، به مفهومی عمومیتر منجر میشود. سایر مزایای بالقوه انتخاب متغیرهای پیش بین شامل تسهیل درک و تجسم دادهها، کاهش الزامات اندازه گیری و ذخیره اطلاعات،



همکاران، ۱۳۹۱، باقرپور و لاشانی و همکاران، ۱۳۹۲، ستایش و همکاران، ۱۳۹۳؛ شبکه‌های بیز ( کرکاس و همکاران، ۲۰۰۷ و ستایش و همکاران، ۱۳۹۳) پرداختند. یافته‌های اغلب این پژوهش‌ها حاکی از برتری روش‌های هوش مصنوعی نسبت به مدل خطی آماری بود.

به‌رغم کاربرد و سودمندی روش‌های تجمیعی در طبقه‌بندی‌های مالی و حسابداری (از قبیل پیش‌بینی بحران مالی)، پژوهشی مشاهده نشد که در آن از این نوع طبقه‌بندی‌کننده‌ها برای پیش‌بینی نوع اظهارنظر حسابرسان استفاده شده باشد. یافته‌های پژوهش‌های بسیاری (از قبیل آلفارو و همکاران، ۲۰۰۸؛ کیم و کانگ، ۲۰۱۰؛ سان و همکاران، ۲۰۱۱؛ راماکریشنان و همکاران، ۲۰۱۵) حاکی از عملکرد بهتر این روش‌های تجمیعی نسبت به روش‌های طبقه‌بندی انفرادی (خطی و سایر روش‌های هوش مصنوعی) در پیش‌بینی بحران مالی بود.

با توجه به مزایای روش‌های تجمیعی نسبت به روش‌های انفرادی و این موضوع که تاکنون پژوهشی مشاهده نشد که در آن به پیش‌بینی اظهارنظر حسابرسان با استفاده از طبقه‌بندی‌کننده تجمیعی لاجیت بوست پرداخته شده باشد، پژوهش حاضر به پیش‌بینی اظهارنظر حسابرسان با استفاده از این نوع طبقه‌بندی‌کننده در بورس اوراق بهادار تهران می‌پردازد. در این راستا، عملکرد این طبقه‌بندی‌کننده‌ها با طبقه‌بندی‌کننده‌های متداول رگرسیون لجستیک و شبکه‌های عصبی مصنوعی مقایسه می‌شود.

#### ۴. فرضیه‌های پژوهش

با توجه به مبانی نظری و پیشینه پژوهش، فرضیه‌های زیر تدوین و آزمون شده است:

۱. تفاوت معناداری بین عملکرد روش تجمیعی لاجیت بوست و روش رگرسیون لجستیک برای پیش‌بینی نوع اظهارنظر حسابرسان وجود دارد.
۲. تفاوت معناداری بین عملکرد روش تجمیعی لاجیت بوست شبکه‌های عصبی مصنوعی برای پیش‌بینی نوع اظهارنظر حسابرسان وجود دارد.
۳. تفاوت معناداری بین عملکرد رگرسیون لجستیک و شبکه‌های عصبی مصنوعی برای پیش‌بینی نوع اظهارنظر حسابرسان وجود دارد.

#### ۵. روش انجام پژوهش

##### ۵.۱. جامعه آماری و نمونه پژوهش

جامعه آماری این پژوهش هشتاد و هشت شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران است که از ابتدای سال ۱۳۸۶ تا پایان ۱۳۹۳ در بورس فعال بوده‌اند. در

این پژوهش از نمونه‌گیری استفاده نشده است، اما نمونه پژوهش براساس معیارهای زیر انتخاب شده است:

۱. جهت افزایش قابلیت مقایسه، سال مالی آنها منتهی به پایان اسفندماه باشد.
۲. به دلیل ماهیت متفاوت فعالیت، جزء شرکتهای سرمایه‌گذاری، بانکها و مؤسسه‌های مالی، نباشد.
۳. اطلاعات مالی مورد نیاز برای انجام این پژوهش را در دوره زمانی مورد بررسی به‌طور کامل ارائه کرده باشند.

با توجه به بررسیهای انجام شده، تعداد ۱۲۱۴ سال- شرکت، حائز شرایط فوق بوده و مورد بررسی قرار گرفته است. از ۱۲۱۴ نمونه مورد بررسی، گزارش حسابرسی ۳۲۷ مورد، مقبول و مابقی تعدیل شده بوده است.

## ۵.۲. طبقه‌بندی‌کننده‌های استفاده شده

در این پژوهش به منظور ارزیابی طبقه‌بندی‌کننده تجمیعی لاجیت بوست، عملکرد پیشینی این طبقه‌بندی‌کننده با عملکرد روش‌های متداول پیشینی (رگرسیون خطی لجستیک و طبقه‌بندی کننده غیرخطی شبکه‌های عصبی)، مقایسه شده است. در ادامه به تشریح این طبقه‌بندی کننده‌ها پرداخته می‌شود.

### ۵.۲.۱. طبقه‌بندی کننده‌های تجمیعی

یکی از طبقه‌بندی کننده‌هایی که در چند سال اخیر مورد توجه قرار گرفته، استفاده از ترکیب چند طبقه‌بندی کننده یا تکرار یک طبقه‌بندی کننده است. این نوع طبقه‌بندی کننده‌ها که با نام طبقه‌بندی کننده‌های تجمیعی شناخته می‌شوند، بر مبنای تصمیم‌گیری چند خبره در مقابل یک خبره عمل می‌کنند. با استفاده از طبقه‌بندی کننده‌های تجمیعی، تصمیم‌های اتخاذ شده، قابل‌تکثیر خواهد بود. لاجیت-بوست یکی از مهمترین الگوریتم‌های بوستینگ<sup>۶</sup> است که با توجه به نحوه عملکرد آن، امروزه کاربردهای زیادی در تخمین و پیشینی متغیرها در زمینهای مختلف پژوهش دارد. این روش در یادگیر بیانظارت مورد استفاده قرار گرفته و از خانواده الگوریتم‌های یادگیری ماشین<sup>۷</sup> به شمار می‌رود. بوستینگ یک فراالگوریتم ترکیبی در حوزه یادگیریماشین است که برای کاهش عدم توازن و همچنین واریانس کار می‌رود. این تکنیک، روشی برای تبدیل سیستم‌های یادگیری ضعیف به قوی بر اساس ترکیب نتایج طبقه‌بندی‌های مختلف است. (Zhou, ۲۰۱۲) بر این اساس، لاجیت-بوست به عنوان یکی از الگوریتم‌های بوستینگ، یادگیریهای ضعیف را به صورت تکرار شونده آموزش داده و به مجموعه قبلی اضافه مینماید تا در نهایت به یک طبقه بندی قوی دست یابد. یادگیرنده‌های ضعیف در حین اضافه شدن به مجموعه، وزن دهی می‌شوند که این وزن دهی معمولاً بر اساس میزان دقت در طبقه‌بندی نمونه‌ها است. وزن دهی نمونه‌ها به صورتی است که در هر مرحله، وزن نمونه‌هایی که به صورت صحیح طبقه‌بندی می‌شوند کاهش یافته و وزن نمونه‌هایی که به درستی طبقه‌بندی نشده‌اند، بیشتر می‌شود تا در مراحل بعدی (توسط یادگیرنده‌های جدید) بیشتر مورد توجه قرار گیرند و با دقت بیشتری طبقه‌بندی شوند؛ بنابراین تمرکز یادگیرنده‌های ضعیف جدید، بیشتر بر روی داده‌هایی خواهد بود که سیستم در مراحل قبلی قادر به طبقه‌بندی صحیح آنها نبوده است. (Smyth, ۲۰۰۱) در واقع، در این الگوریتم با تقسیم مجموعه داده‌های در دسترس به مجموعه‌های آموزش و آزمون و استفاده از الگوریتم یادگیری ماشین، پیشینی متغیرها انجام میشود.

### ۵.۲.۲. رگرسیون لجستیک

<sup>۶</sup>Boosting

<sup>۷</sup>Machine learning

- رگرسیون لجستیک شبیه به رگرسیون معمولی است، با این تفاوت که روش تخمین ضرایب در آنها یکسان نیست. رگرسیون لجستیک، به جای حداقل کردن مجذور خطاها (که در رگرسیون معمولی انجام میشود)، احتمالی که یک واقعه رخ میدهد را حداکثر میکند (مؤمنی و فعالقیومی، ۱۳۸۶: ۱۵۸). در بیشتر پژوهش‌های انجام‌شده در زمینه پیش‌بینی‌گزارش حسابرس سیرایارز یا بیمدل های مبتنی بر هوش مصنوعی، از یک مدل خطی استفاده میشود. با توجه به دلایل زیر، در این پژوهش از مدل خطی رگرسیون لجستیک به عنوان مدل مقایسه‌ای استفاده میشود:

۱. در بین مدل‌های آماری، در پژوهش‌های انجام‌شده در زمینه مالی و حسابداری، از همبستگی استفاده شده و مشهورتر است (راعی و فلاچپور، ۱۳۸۷: ۲۱).
۲. رگرسیون لجستیک نسبت به تحلیل ممیزی (مدل خطی دیگری که در پیش‌بینی گزارش حسابرسی استفاده زیادی از آن می‌شود) ابزار قویتری است (مؤمنی و همکاران، ۱۳۸۶: ۱۵۸).

### ۵.۲.۳. شبکه‌های عصبی مصنوعی

- یک شبکه عصبی مصنوعی، شامل مجموعه‌ای از نرون‌های به هم متصل می‌باشد که به هر مجموعه‌ای از این نرون‌ها یک لایه گفته می‌شود. نقش نرون‌ها در شبکه‌های عصبی، پردازش اطلاعات است. این امر در شبکه‌های عصبی مصنوعی عموماً سیله‌یکپرده از شگرریاضیکه همان تابع فعالساز است، انجام می‌شود. یکتا تابع فعالساز، بر اساس نیاز خاص مسأله‌ای که قرار است به وسیله شبکه عصبی حل شود، از سوی طراح انتخاب می‌شود. ساده‌ترین شبکه، فقط دو لایه دارد. لایه ورودی و لایه خروجی شبکه عصبی سیستم ورودی-خروجی عملی کند و از نرون‌های ورودی برابر با محاسبه ارزش نرون خروجی مورد استفاده قرار می‌دهد (منهاج، ۱۳۷۷: ۱۲۲). در این پژوهش از شبکه‌های عصبی پرسپترون چندلایه با الگوریتم آموزشی پس‌انتشار خطا استفاده شده است. دلایل استفاده از شبکه‌های عصبی پرسپترون چندلایه با الگوریتم آموزشی پس‌انتشار خطا به شرح زیر است:
۱. در دهه‌های اخیر، شبکه‌های عصبی مصنوعی به‌طور وسیعی برای پیش‌بینی‌های مالی استفاده شده است (آلفارو و همکاران، ۲۰۰۸: ۱۱۰، راوی کومار و راوی، ۲۰۰۷). به‌طوری‌که، حدود ۹۵٪ از مطالعات تجاری کاربردی از پرسپترون چندلایه استفاده کرده‌اند و مشهورترین روش آموزشی نیز پس‌انتشار خطا است (تسای، ۲۰۰۹: ۱۲۳).
  ۲. یافته‌های پژوهش‌های تجربی بیانگر دقت بالای این مدل هوش مصنوعی است. این مدل قادر به شناسایی روابط غیرخطی هستند و عملکرد خوبی در شرایط وجود اطلاعات دارای پارازیت از خود نشان می‌دهند (راوی کومار و راوی، ۲۰۰۷؛ آلفارو و همکاران، ۲۰۰۸: ۱۱۰؛ و تسای، ۲۰۰۹: ۱۲۳).

### ۵.۳. روش انتخاب متغیر مبتنی بر همبستگی

- به‌طور کلی یک متغیر مناسب است اگر به متغیر وابسته (طبقه، در طبقه‌بندی) ارتباط داشته‌باشد اما نسبت به سایر متغیرهای پیش‌بینی مربوط، اضافی یا زائد نباشد. اگر همبستگی بین دو متغیر به عنوان معیار مناسب بودن در نظر گرفته شود، تعریف ذکر شده به این صورت خواهد شد که یک متغیر مناسب است در صورتی که همبستگی بالایی با متغیر وابسته (طبقه) و همبستگی پایینی با سایر متغیرهای پیش‌بینی داشته‌باشد. به عبارت دیگر، اگر همبستگی بین یک متغیر پیش‌بینی و متغیر وابسته (طبقه) به اندازه کافی بالا باشد تا برای پیش‌بینی متغیر وابسته (طبقه) مربوط باشد و همبستگی بین آن و سایر متغیرهای پیش‌بینی مربوط، به سطح معنی‌نرسد، به

طوری که نتوان توسط سایر متغیرهای مربوط، متغیر مزبور را پیش‌بینی کرد، آن متغیر به عنوان یک متغیر مناسب برای پیش‌بینی (طبقه‌بندی) قلمداد می‌شود. در این حالت، مساله اصلی انتخاب متغیر، جستجوی معیار مناسب همبستگی بین متغیرها و رویه منطقی برای انتخاب متغیرهای مناسب بر اساس این معیار خواهد بود (یو و لیو<sup>۲۳</sup>، ۲۰۰۳).

روش مبتنی بر همبستگی، همبستگی‌های بین متغیرهای پیش‌بین و همچنین متغیرهای پیش‌بین و متغیر وابسته را محاسبه و سپس فضای زیرمجموعه متغیرها<sup>۲۴</sup> را جستجو می‌کند. زیرمجموعه یافت شده در جستجو که دارای بالاترین سودمندی (براساس رابطه زیر) باشد، برای کاهش ابعاد داده‌های آموزشی اولیه و داده‌های آزمایشی استفاده می‌شود (هال و اسمیت<sup>۲۵</sup>، ۱۹۹۹):

که در آن Merits سودمندی اکتشافی زیرمجموعه متغیر S شامل K متغیر،  $\bar{r}_{cf}$  میانگین همبستگی متغیر مستقل - وابسته و  $\bar{r}_{ff}$  میانگین همبستگی بین متغیر مستقل - متغیر مستقل است. معادله مزبور ضریب همبستگی پیرسون است که در آن تمام متغیرها، استاندارد شده است. صورت کسر، بیانگر کیفیت (حدود) پیش‌بینی کنندگی گروهی از متغیرهای مستقل و مخرج آن بیانگر میزان حشو (اضافی بودن) بین آنهاست.

#### ۴.۵. متغیرهای پژوهش

##### ۴.۵.۱. متغیر وابسته

متغیر وابسته این پژوهش، نوع اظهارنظر حسابرس است. براساس استاندارد حسابرسی بخش ۷۰۰ (گزارشگری نسبت به صورت‌های مالی) اظهارنظر حسابرسان به دو دسته تعدیل نشده (مقبول) و تعدیل شده (شامل اظهارنظرهای مشروط، مردود و عدم اظهارنظر) تقسیم می‌شود. بنابراین، در خصوص سنجش متغیر وابسته از متغیر دو وجهی استفاده شده است که در صورت دریافت گزارش تعدیل شده، عدد یک و در غیر این صورت عدد صفر به هر شرکت اختصاص می‌یابد.

##### ۴.۵.۲. متغیرهای پیش‌بین (مستقل)

در مرحله اول با بررسی ادبیات و پیشینه پژوهش، حدود ۶۰ متغیر پیش‌بین شناسایی شد. از بین متغیرهای شناسایی شده، ۳۵ متغیری که بیشتر در ادبیات پیش‌بینی نوع اظهارنظر حسابرسان استفاده شده و داده‌های مورد نیاز برای سنجش آنها از طریق پایگاه‌های اطلاعاتی سازمان بورس و اوراق بهادار و همچنین نرم‌افزارهای تدبیرپرداز و ره‌آورد نوین در دسترس بود، انتخاب شد. جدول شماره ۱، میانگین این متغیرها و نحوه سنجش آن را نشان می‌دهد. متغیرهای نوع گزارش سال قبل، نوع عملکرد، دعوی حقوقی و اندازه موسسه حسابرسی با استفاده از متغیرهای مجازی، اندازه‌گیری شده است. بدین ترتیب که در صورتی که گزارش حسابرسی سال قبل شرکت، مقبول باشد عدد صفر و در غیر این صورت عدد یک اختصاص یافته است. اگر شرکت سود ده باشد عدد صفر و در غیر این صورت عدد یک اختصاص یافته است. در صورت وجود دعوی حقوقی علیه شرکت عدد یک و در غیر این صورت عدد صفر اختصاص یافته است. در نهایت، در صورتی که حسابرسی شرکت توسط سازمان حسابرسی انجام شده باشد عدد یک و در صورت حسابرسی شرکت توسط سایر موسسات عضو جامعه حسابداران رسمی، عدد صفر اختصاص یافته است. در ادامه، با استفاده از روش انتخاب متغیرهای پیش‌بین مبتنی بر همبستگی در نرم‌افزار Weka نسخه ۷-۳ از بین ۳۵ متغیر ذکر شده، متغیرهای بهینه، انتخاب شده است. اجرای روش انتخاب متغیر مبتنی بر همبستگی منجر به انتخاب ۶ متغیر از بین ۳۵ متغیر اولیه شد



(متغیرهای ردیف‌های ۱، ۳، ۶، ۲۰، ۲۹ و ۳۲ در جدول شماره ۱). شایان ذکر است که در این پژوهش، از داده‌های یک سال قبل شرکتها برای پیشبینی نوع اظهار نظر حسابرس استفاده شده است.

جدول ۱- متغیرهای پیش‌بین و نحوه سنجش آن

ردیف	متغیرهای پیش‌بین	نحوه سنجش	میانگین گزارش‌های مقبول	میانگین گزارش‌های تعدیل شده
۱	نوع گزارش حسابرسی سال قبل			
۲	نوع عملکرد			
۳	دعای حقوقی			
۴	اندازه موسسه حسابرسی			
۵	سن (عمر) شرکت بر حسب سال	LD	۲۱/۴۲	۱۹/۳۴
۶	حاشیه سود ناخالص	GP/S	۰/۳۲	۰/۲۳
۷	حاشیه سود خالص	NI/S	۰/۲۷	۰/۱۰
۸	بازده حقوق صاحبان سهام	NI/SE	۰/۵۱	۰/۳۹
۹	بازده کل داراییها	NI/TA	۰/۱۵	۰/۱۱
۱۰	سود انباشته به مجموع داراییها	RE/TA	۰/۱۹	۰/۱۰
۱۱	نسبت آبی	(CA-Inv-PP)/CL	۰/۷۶	۰/۶۲
۱۲	نسبت جاری	CA/CL	۱/۱۸	۱/۰۷
۱۳	نسبت سرمایه در گردش به داراییها	(CA-CL)/TA	۰/۱۴	۰/۱۱
۱۴	نسبت حسابهای دریافتی به کل داراییها	AR/TA	۰/۲۴	۰/۲۷
۱۵	لگاریتم طبیعی کل داراییها (اندازه شرکت)	Ln (TA)	۱۳/۲	۱۲/۸
۱۶	لگاریتم طبیعی فروش خالص	Ln (S)	۱۲/۹	۱۲/۷
۱۷	لگاریتم طبیعی تعداد کارکنان	Ln (PN)	۶/۱۲	۵/۹۷
۱۸	نسبت بدهی (اهرم مالی)	TL/TA	۰/۶۱	۰/۶۹
۱۹	نسبت بدهی بلندمدت به حقوق صاحبان سهام	LTD/SE	۰/۲۷	۰/۳۳
۲۰	پوشش هزینه‌های مالی	EBIT/IE	۹۸/۵۴	۹۳/۳۱
۲۱	سرنانه ذخیره مزایای پایان خدمت کارکنان	RP/PN	۲۷/۳۴	۲۴/۴۵
۲۲	سرنانه داراییها	TA/PN	۹۲۱/۲۳	۸۷۵/۴۵
۲۳	سرنانه فروش	S/PN	۷۷۴/۸۵	۶۵۷/۷۶
۲۴	سرنانه سودخالص	NI/PN	۱۵۶/۳۷	۷۵/۶۴
۲۵	دوره وصول مطالبات	ART/۳۶۵	۱۳۱/۴۶	۱۴۲/۳۷
۲۶	گردش متوسط مجموع داراییها	S/TA <sub>(ave)</sub>	۰/۷۵	۰/۸۱
۲۷	گردش متوسط داراییهای ثابت	S/FA <sub>(ave)</sub>	۶/۵۸	۷/۶۲
۲۸	گردش متوسط حسابهای دریافتی	S/AR <sub>(ave)</sub>	۴/۳۸	۴/۱۴
۲۹	گردش متوسط موجودی کالا	COGS/Inv <sub>(ave)</sub>	۱۰/۴۶	۴/۹۳
۳۰	وجه نقد ناشی از فعالیتهای عملیاتی به فروش	OCF/S	۰/۱۷	۰/۱۲
۳۱	وجه نقد فعالیتهای سرمایه‌گذاری به فروش	ICF/S	-۰/۱۲	-۰/۰۷
۳۲	ذخیره مالیات به فروش	TR/S	۰/۰۳	۰/۰۴
۳۳	سود (زیان) انباشته به فروش	RE/S	۰/۱۷	۰/۰۷
۳۴	رشد سود	%Δ (NI)	۰/۲۶	۰/۱۸
۳۵	رشد شرکت	%Δ (TA)	۰/۲۸	۰/۲۰

LD: تاریخ درج GP: سود ناخالص S: فروش NI: سود خالص SE: حقوق صاحبان سهام TA: مجموع داراییها RE: سود انباشته CA: داراییهای جاری Inv: موجودی کالا PP: پیش پرداختها CL: بدهیهای جاری AR: حسابها و اسناد

در تلفتی PN: تعداد پرسنل TL: مجموع بدهی ها LTD: بدهی های بلندمدت EBIT: سود قبل بهره و مالیات IE: هزینه های مالی RP: ذخیره مزایای پان خدمت ART: گردش مطالبات FA: دارایی های ثابت COGS: بهای تمام شده کالا ی فروش رفته OCF: خالص جریان نقد عملیاتی ICF: خالص جریان نقدی فعالیت های سرمای گذاری TR: ذخیره مالیت %: درصد رشد سالانه

## ۵.۵. روایی متقابل

در روش holdout که در اغلب پژوهش های حسابداری و مالی (به ویژه در ایران) استفاده شده است، داده ها به دو دسته به نام مجموعه آموزشی و مجموعه آزمائشی تقسیم می شود. این روش ها دارای محدودیت های بارزی هستند. روش holdout تخمین گر بدبختانه است، زیرا تنها بخشی از داده ها برای آموزش به روش پیش بینی ارائه شده است. هر چه تعداد نمونه بیشتری برای مجموعه آزمائشی خارج شود، تعصب برآورد بیشتر می شود. از طرفی، نمونه های آزمائشی کوچک تر (با تعداد کمتر) به معنای این است که فاصله اطمینان دقت، بیشتر خواهد بود. بنابراین، روش مزبور، روش مناسبی نخواهد بود (کوهاوی، ۱۹۹۵). در مقابل، روش روایی متقابل، به دلیل سادگی، شفافیت و جامعیت، یک نگاه برد متداول و مناسب است و نتایج بسطی از پژوهش های انجام شده حاکی از عملکرد بهتر این روش است. در این راستا، در پژوهش حاضر به منظور بررسی تعمیم پذیری پیش بینی های انجام شده از روایی متقابل ۱۰ بخش استفاده می شود. روایی متقابل ۱۰ بخشی برای برآورد نرخ خطای واقعی کاملاً قابل اتکا و کافی است (هو، ۲۰۱۰). در این روش، نمونه اصلی به ۱۰ دسته نمونه فرعی مختلف تقسیم می شود. ۹ نمونه فرعی به عنوان نمونه های آموزشی استفاده می شود و نمونه فرعی باقیمانده به عنوان نمونه آزمائشی، مورد آزمون قرار می گیرد. این شیوه تا حدی تکرار می شود که هر یک از ۱۰ نمونه فرعی به عنوان نمونه آزمائشی مورد آزمون قرار گیرد. در این پژوهش، روایی متقابل ۱۰ بخشی، با استفاده از اجزای مختلف مجموعه داده ها، به طور مستقل، ۱۰ بار انجام شده است (روایی متقابل ۱۰ بخشی با ۱۰ بار تکرار). یافته های اغلب پژوهش ها (به عنوان نمونه، کوهاوی، ۱۹۹۵) حاکی از آن است که در مسائل دنیای واقعی، روایی متقابل ۱۰ بخشی، بهترین روش انتخاب مدل است. مزیت روش مزبور، این است که تمام نمونه ها در نهایت هم به عنوان داده های آموزشی و هم به عنوان داده های آزمائشی استفاده خواهد شد. افزون بر این، استفاده از روایی متقابل، از بروز مشکل بیش برآزش و مشکلات مربوط به نتایج برون نمونه ای جلوگیری می کند.

## ۵.۵. روش آزمون فرضیه ها

به منظور ارزیابی عملکرد پیش بینی، از معیارهای ارزیابی (شامل میانگین دقت و خطاهای نوع اول و دوم) استفاده می شود. به عبارت دیگر، میانگین معیارهای ارزیابی مربوط به پیش بینی نوع اظهار نظر حسابرسان با استفاده از روش تجمیعی لاجیت بوست با معیارهای ارزیابی روش متداول خطی رگرسیون لجستیک و روش متداول غیرخطی شبکه های عصبی مقایسه می شود. پیش بینی نوع اظهار نظر به وسیله نرم افزار Weka نسخه ۳-۷ انجام شده است. به منظور آزمون فرضیه های پژوهش نیز براساس صد دقت حاصل از اجرای روایی متقابل ۱۰ بخش با ۱۰ بار تکرار، از آزمون t زوجی یا آزمون ناپارامتریک معادل آن استفاده شده است. در این خصوص از نرم افزار SPSS نسخه ۲۱ استفاده گردید.

## ۶. یافته‌های تجربی پژوهش

جدول شماره ۲، میانگین دقت، خطای نوع اول و دوم مربوط به پیش‌بینی نوع اظهار نظر حسابرسان را براساس طبقه‌بندی‌کننده‌های لاجیت بوست، رگرسیون لجستیک و شبکه‌های عصبی با استفاده از ۶ متغیر پیش‌بین انتخاب شده، نشان می‌دهد. شایان ذکر است که میانگین معیارها مربوط به صد معیار حاصل از اجرای رویی متقابل ۱۰ بخشی با ۱۰ بار تکرار است.

جدول ۲: عملکرد طبقه‌بندی‌کننده‌های لاجیت بوست، لجستیک و شبکه عصبی

طبقه‌بندی‌کننده	عملکرد	لاجیت بوست	لجستیک	شبکه عصبی
دقت	۹۴/۳۴	۸۵/۲۱	۹۲/۳۷	
خطای نوع اول	۰/۰۳	۰/۱۴	۰/۰۵	
خطای نوع دوم	۰/۰۹	۰/۱۷	۰/۱۰	

نتایج بررسی نرمال بودن صد دقت در هر یک از طبقه‌بندی‌کننده‌ها حاکی از نرمال بودن آن بود. جدول شماره ۳ نتایج آزمون آزوجی و مقدار احتمال مربوطه را برای مقایسه میانگین دقت هر جفت طبقه‌بندی‌کننده نشان می‌دهد. با توجه به آماره  $t$  و مقدار احتمال مربوطه ( $p < 0/05$ )، تفاوت معناداری بین میانگین دقت طبقه‌بندی‌کننده‌های لاجیت بوست و رگرسیون لجستیک وجود دارد. بنابراین، فرضیه اول پژوهش، تأیید می‌شود. در نتیجه، با توجه به بالاتر بودن دقت طبقه‌بندی‌کننده لاجیت بوست می‌توان چنین استنباط کرد که روش تجمیعی لاجیت بوست به طور معناداری بهتر از رگرسیون لجستیک، نوع اظهار نظر حسابرسان را پیش‌بینی می‌کند. افزون بر این، با توجه به آماره  $t$  و مقدار احتمال مربوطه ( $p < 0/05$ )، تفاوت معناداری بین میانگین دقت طبقه‌بندی‌کننده‌های لاجیت بوست و شبکه‌های عصبی وجود دارد. بنابراین، فرضیه دوم پژوهش، تأیید می‌شود. در نتیجه، با توجه به بالاتر بودن دقت طبقه‌بندی‌کننده لاجیت بوست می‌توان چنین استنباط کرد که روش تجمیعی لاجیت بوست به طور معناداری بهتر از شبکه‌های عصبی، نوع اظهار نظر حسابرسان را پیش‌بینی می‌کند. در نهایت، با توجه به آماره  $t$  و مقدار احتمال مربوطه ( $p < 0/05$ )، تفاوت معناداری بین میانگین دقت طبقه‌بندی‌کننده‌های لجستیک و شبکه‌های عصبی وجود دارد. بنابراین، فرضیه سوم پژوهش نیز تأیید می‌شود. در نتیجه، با توجه به بالاتر بودن دقت طبقه‌بندی‌کننده‌های لاجیت بوست می‌توان چنین استنباط کرد که این روش تجمیعی به طور معناداری بهتر از شبکه‌های عصبی، نوع اظهار نظر حسابرسان را پیش‌بینی می‌کند. در مورد میانگین خطای نوع اول و دوم نیز آزمون  $t$  زوجی انجام و نتایج فوق یکسان بود که ارائه جداول و تفسیرهای مربوطه خارج از حوصله این مقاله است.

جدول ۳: نتایج آزمون آزوجی و مقدار احتمال مربوط به مقایسه دقت هر جفت از طبقه‌بندی‌کننده‌ها

طبقه‌بندی‌کننده	لاجیت بوست	لجستیک	شبکه عصبی
لاجیت بوست			

		۹/۵۴۶ (۰/۰۰۰)	لجستیک
	۷/۳۶۸ (۰/۰۰۰)	۴/۶۴۵ (۰/۰۰۰)	شبکه عصبی

## ۷. خلاصه و نتیجه‌گیری

با توجه به محتوای اطلاعاتی گزارش حسابرس مستقل، پیشبینی نوع اظهارنظر در تصمیم‌گیریهای مالی از اهمیت بسزایی برخوردار است. بهره‌مندی دقت و عملکرد بهتر طبقه‌بندی‌کننده‌های تجمیعی نسبت به تکنیکهای آماری و بسیاری از تکنیکهای هوش مصنوعی در پیشبینی، تاکنون پژوهشی با استفاده از این طبقه‌بندی‌کننده‌ها در بورس اوراق بهادار تهران انجام نشده است. بنابراین، در پژوهش حاضر به ارزیابی طبقه‌بندی‌کننده تجمیعی لاجیت بوست برای پیشبینی نوع اظهارنظر حسابرسان در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران پرداخته شد. افزون بر این، وجود تعداد متغیرهای زیاد، نه تنها بر عملکرد پیشبینی‌کننده، اثر نامساعد می‌گذارد، بلکه بر زمان اجرای الگوریتم یادگیری را نیز تحت تاثیر قرار می‌دهد. هرچه تعداد متغیرها بیشتر شود، زمان اجرای الگوریتم یادگیری نیز بیشتر می‌شود. همچنین بعد بالای داده‌ها می‌تواند به مساله اضافه بار ابعاد منجر شود. بنابراین، در این پژوهش به منظور انتخاب متغیرهای پیشین بهینه از روش انتخاب متغیر مبتنی بر همبستگی استفاده شد که منجر به انتخاب ۶ متغیر بهینه از بین متغیرهای اولیه گردید. یافته‌های تجربی پژوهش حاکی از آن است که طبقه‌بندی‌کننده تجمیعی لاجیت بوست، به‌طور معناداری، بهتر از رگرسیون لجستیک و شبکه‌های عصبی، نوع اظهارنظر حسابرسان را پیشبینی می‌کنند. همچنین، شبکه‌های عصبی به طور معناداری بهتر از رگرسیون لجستیک، نوع اظهارنظر حسابرسان را پیش‌بینی می‌کند.

## منابع و مآخذ

- احمدپور، احمد؛ طاهرآبادی، علی اصغر؛ و شعیعی‌عباسی (۱۳۸۹). «تاثیر متغیرهای مالی و غیرمالی بر صدور اظهارنظر مشروط حسابرسی (شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران)». فصلنامه مهور ساورا قبهادار تهران، شماره ۹، صفحات ۹۷-۱۱۴.
- ارباب سلیمانی، عباس و محمود نفری (۱۳۸۵). اصول حسابرسی. جلد اول، چاپ شانزدهم، تهران: مرکز پژوهشات تخصصی حسابداری و حسابرسی - سازمان حسابرسی.
- امیر اصلانی، حامی (۱۳۸۴). نقش اقتصادی حسابرسی در بازارهای آزاد و بازارهای تحت نظارت. تهران: مدیریت تدوین استانداردها - سازمان حسابرسی.
- امینی، پیمان؛ محمدی، کامران؛ و شعیعی‌عباسی (۱۳۹۰). «بررسی عوامل موثر بر صدور گزارش مشروط حسابرسی: کاربرد روش شبکه عصبی». حسابداری مدی‌ت، سال چهارم، شماره هیازدهم، صفحات ۳۹-۲۵.
- باقرپور و لاشانی، محمد علی؛ ساعی، محمد جواد؛ مشکانی، علی و مصطفی باقری (۱۳۹۲). پیش‌بینی گزارش حسابرس مستقل در ایران: رویکرد داده کاوی، فصلنامه تحقیقات حسابداری و حسابرسی، سال پنجم، شماره ۱۹، صص. ۱۳۴-۱۵۰.
- بنیمهد، بهمن و سعید علیاحمدی (۱۳۹۲). بررسی تحلیلی سودمند یگزارش‌های حسابرس سید بازار سرمایه، دانش حسابداری و حسابرس سیمدیریت، دوره ۲، شماره ۶، صص. ۱۳-۲۹.



- DeAngelo, L. (1981), 'Auditor Size and Auditor Quality', *Journal of Accounting and Economics*, Vol. 1, pp. 113–27.
- Doumpos, M.; Gaganis, C.; and F. Pasiouras (2005). Explaining qualifications in audit reports using a support vector machine methodology. *Intelligent Systems in Accounting, Finance and Management*, Volume 13, Issue 4, pages 197–215.
- Firth, M (1980). "A note on the impact of audit qualifications on lending and credit decisions". *Journal of Banking & Finance*, Vol. 4 No.3, pp. 257-267.
- Freund, Y., and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55 (1): 119–139.
- Gaganis Ch., Pasiouras F., Doumpos M., (2007), "Probabilistic Neural Networks for the Identification of Qualified Audit Opinions", *Expert Systems with Applications*, vol. 32, no.1, pp. 114-124.
- Gaganis, C., F. Pasiouras, C. Spathis, and Zopounidis, C., (2007), "A Comparison of Nearest Neighbours, Discriminant and Logit Models for Auditing Decisions", *Intelligent Systems in Accounting, Finance and Management*, Vol. 15, no. 1-2, pp. 23-40.
- Hall, M. A. (2000) "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning", In *Proceedings of the Seventeenth international Conference on Machine Learning (June 29 - July 02)*, pp. 359-366.
- Hu, Y. C. (2010). "Analytic Network Process for Pattern Classification Problems Using Genetic Algorithms", *Information Sciences*, Vol. 180, pp. 2528–2539.
- Hung, C. and Chen. J (2009). A Selective Ensemble Based on Expected Probabilities for Bankruptcy Prediction. *Expert Systems with Applications*, 36: 5297–5303.
- Hung, C. and Chen. J (2009). A Selective Ensemble Based on Expected Probabilities for Bankruptcy Prediction. *Expert Systems with Applications*, 36: 5297–5303.
- Keasey K, Watson R, Wynarczyk P., (1987), 'The small company audit qualification: a preliminary investigation', *Accounting and Business Research*, Vol 18, PP. 323–333.
- Kim, M. and Kang, D. (2010). Ensemble with Neural Networks for Bankruptcy Prediction. *Expert Systems with Applications*, 37: 3373–3379.
- Kirkos, E.; Spathis, C.; Nanopoulos, A.; and Manolopoulos, Y (2007). "Identifying Qualified Auditors opinion: A Data Mining Approach", *Journal of Emerging technologies in Accounting*, Vol. 4, pp. 183-197.
- Kohavi, R. (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", *IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence*, pp.1137-1143.
- Lindenbaum, M., Markovitch, S., and Rusakov, D. (2004). "Selective Sampling for Nearest Neighbor Classifiers", *Machine Learning*, Vol. 2, pp. 125-152.
- Ramakrishnan, S., Mirzaei, M., and Bekri, M. (2015). Adaboost Ensemble Classifiers for Corporate Default Prediction. *Research Journal of Applied Sciences, Engineering and Technology*, 9(3): 224-230.
- Ravi Kumar, P. and Ravi, V (2007). "Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review." *European Journal of Operational Research*, Vol. 180, pp. 1-28.
- Sun, J., Jia, M., and Li, H. (2011). AdaBoost Ensemble for Financial Distress Prediction: An Empirical Comparison with Data from Chinese Listed Companies. *Expert Systems with Applications*, 38 (8): 9305–9312.
- Tsai, C. (2009). Feature Selection in Bankruptcy Prediction. *Knowledge-Based Systems*, 22(2): 120–127.
- Hand, D., Mannila, H and Smyth, P. (2001). *Principles of Data Mining*. Massachusetts: MIT Press.
- Zhou, Z.-H. (2012). *Ensemble Methods, Foundations and Algorithms*. New York: CRC Press.