فصلنامه علمی پژوهشی
دانش حسابداری و حسابرسی مدیریت
دوره ۱۶/ شماره ۳ (پیاپی ۶۳)/ پائیز ۱۴۰۶
صفحه ۱۷۵ تا ۱۹۱

بررسی کارائی و دقت الگوریتم های یادگیری ماشین در پیش بینی نوع اظهارنظر حسابرسی: شواهدی از بورس اوراق بهادار تهران

علیرضا رحیم زاده
گروه حسابداری، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران
rahimzadeh.al@gmail.com

مهران متین فرد
گروه حسابداری، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران (نویسنده مسئول)
mehraan.matinfard@gmail.com

زهره حاجیها
گروه حسابداری، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران
drzhajiha@gmail.com

احسان رحمانی نیا
گروه حسابداری، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران
ehsanrahmaninia@gmail.com

تاریخ دریافت: ۱۴۰۳/۰۴/۰۸ تاریخ پذیرش: ۱۴۰۳/۰۸/۱۴

چکیده

پژوهش حاضر به بررسی و مقایسه دقت و کارائی الگوریتم های یادگیری ماشین جهت پیش بینی نوع اظهارنظر حسابرسی در شرکت های پذیرفته شده در بورس اوراق بهادار تهران می پردازد. در این راستا از روش انتخاب متغیر (آزمون مقایسه میانگین دو نمونه) جهت بررسی و انتخاب متغیرهای تاثیر گذار بر نوع اظهار نظر حسابرسی استفاده شده است. به منظور نیل به این هدف ۱۶۰۶ سال-شرکت (۱۴۶ شرکت برای ۱۱ سال) مشاهده جمع آوری شده از گزارشات مالی سالیانه شرکت های پذیرفته شده در بورس اوراق بهادار تهران در طی دوره زمانی ۱۳۹۰ تا ۱۴۰۰ مورد آزمون قرار گرفته اند. در پژوهش حاضر از شش الگوریتم یادگیری ماشین (درخت تصمیم و رگرسیون، جنگل تصادفی، شبکه عصبی، نزدیکترین همسایگی، رگرسیون لوجیت، ماشین بردار پشتیبان) و همچنین یافته های حاصل از تکنیک های یادگیری ماشین نشان می دهد که صحت کلی تکنیک های درخت تصمیم و رگرسیون، جنگل تصادفی، شبکه عصبی، نزدیکترین همسایگی، رگرسیون لوجیت، ماشین بردار پشتیبان به ترتیب ۷۶.۳٪، ۷۷.۷٪، ۷۶.۹٪، ۷۴.۶٪، ۷۸.۳٪ و ۶۹.۶٪ می باشد که نشان دهنده کارآ بودن الگوریتم رگرسیون لوجیت نسبت به سایر الگوریتم های یادگیری ماشین می باشد. در ضمن در حالت کلی میزان خطای مدل در شناسایی شرکت های دارای اظهارنظر حسابرسی مقبول با الگوریتم درخت تصمیم و رگرسیون ۰.۲۵۱ است و برای الگوریتمهای جنگل تصادفی، شبکه عصبی، نزدیکترین همسایگی، رگرسیون لوجیت، ماشین بردار پشتیبان به ترتیب ۰.۰۹۳، ۰.۱۲۱، ۰.۱۶۷، ۰.۱۰۲، ۰.۰۸۲ می باشد. بنابراین با توجه به روش انتخاب متغیرها ، الگوریتم ماشین بردار پشتیبان (SVM) دارای کمترین میزان خطا و در واقع بیشترین دقت را دارد.

واژه‌های کلیدی: نوع اظهارنظر حسابرسی ، پیش بینی، الگوریتم یادگیری ماشین، آزمون مقایسه میانگین دو نمونه.

۱- مقدمه

انجام حسابرسی به عنوان یک فعالیت نظارتی و نیز به عنوان یکی از مکانیزم های حاکمیت شرکتی برای جلب (و حفظ) اعتماد سرمایه گذاران به وجود تقارن اطلاعاتی، منصفانه بودن بازار و نیز منصفانه بودن اطلاعات در دسترس ضروری است. انجام حسابرسی، به شرط برخورداری از سطح کیفی مناسب، نیاز استفاده کنندگان گزارش های مالی به ارزیابی کیفیت اطلاعات را پیش از استفاده از آنها در فرآیند تصمیم گیری برطرف می سازد و زمینه مناسبی برای اخذ تصمیمات اقتصادی مطلوب فراهم می آورد (انجمن حسابداری آمریکا، ۱۹۷۳). صورت های مالی حسابرسی شده به وسیله حسابرسان مستقل، وسیله بسیار مناسبی در انتقال اطلاعات قابل اتکاست. حسابرس مستقل، شایسته ترین شخص برای اظهار نظر درباره قابلیت اعتماد گزارش های مالی واحد اقتصادی است . شایستگی حسابرس به این دلیل است که وی فردی مستقل بوده و حسابرسی را طبق استانداردهای حسابرسی انجام می دهد تا اطمینان یابد که اقلام صور ت های مالی مطابق با استانداردهای حسابداری تهیه شده است. بنابراین، حسابرس به ادعاهایی اعتبار می دهد که توسط شخص دیگری در قالب صورت های مالی تهیه شده است و بدین وسیله قابلیت اتکای اطلاعات به کار رفته در تصمیمات اقتصادی را افزایش می دهد (نیکخواه آزاد،۱۳۷۹).

نقش حسابرسان برای ارائه تضمین کافی و مناسب برای کنترل مباحث اقتصادی شرکتها در حال حاضر تکامل یافته است. از طرفی حسابرسی از ارکان اساسی فرآیند پاسخ گویی است، زیرا پاسخ گویی مستلزم وجود اطلاعات معتبر و قابل اتکاست و قابلیت اتکای اطلاعات، مستلزم بررسی آن ها به دست شخصی مستقل از تهیه کننده اطلاعات است. این امر مهم از طریق فرآیند حسابرسی صورت می گیرد که در فرآیند پاسخ گویی، حسابرسی با اظهار نظر مربوط از طریق تعیین اعتبار اطلاعات، ارزش افزوده ایجاد می کند. بدین صورت، یک پل ارتباطی بین یافته های حسابرسان با افراد خارج و داخل شرکت به صورت اظهار نظر حسابرسی به وجود می آید که نقش مهمی در هشدار به استفاده کنندگان از صورت های مالی نسبت به شناخت مشکلات پیش روی شرکت ایفا می کند و اگر شرکت دارای هر گونه ایراد و تحریفی باشد، موسسه حسابرسی به سهولت از آن نخواهد گذشت و بر روی اظهار نظر حسابرسی تأثیر گذار خواهد بود(دی آنجلو، ۱۹۸۱).

الگوریتم های یادگیری ماشینی یا داده کاوی که هدف آن استخراج اطلاعات موردعلاقه از مجموعه گسترده ای از داده ها است به طور گسترده ای به عنوان یک ابزار تصمیم گیری فعال مورداستفاده قرارگرفته است. این امر باعث افزایش توجه پژوهشگران در غنی کردن داده ها باهدف تسهیل کشف الگوهای موردعلاقه شده است (وایتینگ و همکاران، ۲۰۱۲). داده کاوی، استخراج یا کاوش دانش از حجم زیادی اطلاعات است. الگوها یا قواعد قوی کشف شده از طریق تکنیک های داده کاوی را می توان برای پیش بینی غیر بدیهی داده های جدید استفاده کرد. ازاین رو، داده کاوی یک حوزه میان رشته ای است که از ابزارهای تجزیه و تحلیل مدل های آماری، الگوریتم های ریاضیاتی و روش های یادگیری ماشینی برای کشف الگوها و روابط معتبر قبلاً ناشناخته در مجموعه داده های بزرگ استفاده می کند(دوا و دو، ۲۰۱۱). با توجه به مطالب بیان شده، هدف پژوهش حاضر بررسی کارائی و دقت الگوریتم های یادگیری ماشین در پیش بینی نوع اظهارنظر حسابرسی در شرکت های پذیرفته شده در بورس اوراق بهادار تهران می باشد. جهت دستیابی به هدف مذکور، از اطلاعات ۱۴۶ شرکت در طی دوره زمانی ۱۳۹۰ تا ۱۴۰۰ استفاده است.

۲. مبانی نظری و پیشینه پژوهش

الگوریتم‌های یادگیری ماشین که هدف اصلی آن‌ها استخراج و تحلیل اطلاعات از حجم گسترده‌ای از داده‌ها است، به‌عنوان ابزارهای تصمیم‌گیری پیشرفته در حوزه‌های مختلف از جمله حسابرسی مورد استفاده قرار می‌گیرند. این الگوریتم‌ها، با کشف الگوها و روابط پنهان در داده‌ها، می‌توانند به‌عنوان ابزاری قدرتمند در پیش‌بینی اظهارنظر حسابرسی و ارتقای دقت و کیفیت گزارش‌های حسابرسی به کار گرفته شوند.

مشکلات نمایندگی بین سهامداران و مدیران منجر به بکارگیری حسابرسان می شود، که اطمینان بخشی مستقلی برای سهامداران فراهم می کند مبنی بر اینکه صورت های مالی تهیه شده توسط مدیران شرکت مطابق با اصول عمومی پذیرفته شده حسابداری می باشد(واتز و زیمرمن ، ۱۹۸۳). علاوه بر این حسابرسی از طریق کشف سلب مالکیت توسط افراد درون سازمانی(نئومن و همکاران ، ۲۰۰۵)، و منتفع کردن مدیریت شرکت از طریق پیام دهی در خصوص قابلیت اعتماد اطلاعات مالی تهیه شده توسط مدیریت، نقش مهمی در اجرا قانون و حمایت از حقوق سرمایه گذاران ایفا می کند(حبیب و همکاران ، ۲۰۱۷).

اظهارنظر حسابرسی، محصول نهایی فرآیند حسابرسی است و کیفیت آن به‌عنوان نشانه‌ای از کیفیت صورت‌های مالی شرکت‌ها مورد ارزیابی قرار می‌گیرد. گزارش‌های تعدیلی حسابرسان معمولاً نشان‌دهنده مشکلاتی در صورت‌های مالی هستند که می‌تواند منجر به کاهش کیفیت سود شرکت شود. این موضوع اهمیت استقلال حسابرس را برجسته می‌کند، زیرا

عدم استقلال ممکن است اعتبار گزارش‌های مالی را مخدوش کرده و تصمیم‌گیری‌های نادرست توسط استفاده‌کنندگان اطلاعات را به دنبال داشته باشد (دی‌آنجلو، ۱۹۸۱).

عوامل مختلفی می‌تواند بر اظهارنظر حسابرسی تأثیرگذار باشد. به رغم اهمیت این موضوع، تحقیقات جامعی برای پیش‌بینی نوع اظهارنظر حسابرسی در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران انجام نشده است. لذا، هدف این پژوهش ارائه الگویی مبتنی بر الگوریتم‌های یادگیری ماشین برای پیش‌بینی نوع اظهارنظر حسابرسی در این شرکت‌ها است.

تکنیک‌های یادگیری ماشین مورد استفاده در پژوهش

در ادامه به پیروی از پژوهش‌های استانیشیچ و همکاران (۲۰۱۹)، البرشی و همکاران (۲۰۲۳)، و هانت و همکاران (۲۰۲۱)، در این پژوهش به منظور کشف نوع اظهارنظر حسابرسی از الگوریتم‌های یادگیری ماشین زیر استفاده خواهد شد. در نهایت، هر یک از این الگوریتم‌ها که قدرت پیش‌بینی بیشتری داشته باشند، به منظور ارائه استفاده خواهند شد:

- درخت تصمیم و رگرسیون(CART)
- جنگل تصادفی(RF)
- شبکه عصبی(KNN)
- نزدیکترین همسایگی(ANN)
- رگرسیون لوجیت(LR)
- ماشین بردار پشتیبان (SVM)

پیشینه داخلی

میرزائی و همکاران (۱۴۰۲) در پژوهشی به این نتیجه رسیدند که مقایسه عملکرد مدل‌های یادگیری ماشین و آماری در پیش‌بینی سود خالص و جریان نقد عملیاتی آتی نشان می‌دهد که رویکرد هوش مصنوعی توانایی بیشتری دارد و بین مدل‌های یادگیری ماشین، رگرسیون نمادین و مدل‌های آماری، مدل پروبیت از عملکرد بیشتری برخوردار است. همچنین نتایج نشان داد که اگرچه به‌طور میانگین مدل‌های یادگیری ماشین عملکرد بهتری نسبت به مدل‌های آماری دارند، مدل‌های آماری نیز عملکرد بیشتری از برخی مدل‌های یادگیری ماشین ارائه می‌دهند.

هاشمی و همکاران (۱۴۰۱) در پژوهشی کاربرد یادگیری ماشین در ارائه الگویی برای پیش‌بینی ورشکستگی را بررسی کرده‌اند. نتایج پژوهش نشان می‌دهد از میان نسبت‌های مالی مشخص شده در مدل اول، تنها نسبت درآمد خالص به کل دارایی و نسبت ارزش بازار حقوق صاحبان سهام به کل ارزش بازار می‌تواند توانایی مدل پیش‌بینی ورشکستگی را بهبود ببخشد. همچنین، در مدل دوم، نسبت‌های مالی مشخص شده توانایی بهبود مدل پیش‌بینی ورشکستگی را داشته و با اضافه کردن متغیر Devscore برای گروه‌هایی که بر اساس صنعت و اندازه ساخته شده‌اند، مدل اصلاح شده و پیش‌بینی ورشکستگی را بهبود می‌بخشد. نتایج حاکی از آن است که یک شرکت در صورت داشتن نسبت‌های مالی مرتبط با ورشکستگی که پایین‌تر از میانگین همسانان خوشه‌ای خود باشد، احتمالاً ورشکسته می‌شود.

فیضی زاده (۱۴۰۰) در پژوهشی تأثیر شواهد حسابرسی بر گزارش حسابرس را مورد بررسی قرار داد. یافته‌های تجربی حاصل از رگرسیون لاجستیک دوتایی نشان داد که کفایت شواهد حسابرسی منفی اما ناچیز است و اعتبار شواهد حسابرسی دارای نشانه ضریب مثبت، اما ناچیز است. این مطالعه نشان داد که فعالیت‌های تجربی بیشتری باید در این زمینه انجام شود.

فروزنده و همکاران (۱۴۰۰) در پژوهشی به این نتیجه رسیدند که کیفیت پایین صورت‌های مالی بر اساس معیار شدت تجدید ارائه بر نوع گزارش حسابرسی، افزایش تعداد بندهای شرط، بندهای تأکید بر مطلب خاص، و سایر بندهای توضیحی اثری ندارد. حسابرسان، طبق نظریه توجه محدود، برای صورت‌های مالی حاوی تحریف‌ها و اشتباهات زیاد که در سال بعد با شدت بالا تجدید ارائه شده‌اند، گزارش حسابرسی تعدیل شده با بندهای بیشتر شرط و توضیحی صادر نکرده‌اند تا استفاده‌کننده از صورت‌های مالی از کیفیت پایین آن آگاه شود. لذا قابلیت اتکاء و اعتماد به گزارش‌های حسابرسی بر اساس معیار شدت تجدید ارائه پایین است.

امری و همکاران (۱۴۰۰) در پژوهشی به این نتیجه رسیدند که کوته‌بینی و توانایی مدیران با اظهارنظر حسابرس و تغییر حسابرس در سال بعد از اظهارنظر تعدیل‌شده ارتباطی ندارد. توانایی مدیریت نتوانسته است ارتباط بین کوته‌بینی مدیریت و صدور گزارش حسابرسی تعدیل‌شده و ارتباط بین کوته‌بینی مدیریت و احتمال تغییر حسابرس پس از صدور اظهارنظر تعدیل‌شده را تعدیل کند. از دلایل عدم ارتباط مزبور می‌توان به مشکل بودن شناسایی دستکاری‌های واقعی سود توسط حسابرسان به علت ماهیت آن‌ها و عدم تأثیرپذیری حسابرسان نسبت به توانایی مدیریت، متفاوت بودن بازار حسابرسی ایران نسبت به دیگر کشورها و غیره اشاره کرد.

خواجوی و همکاران (۱۳۹۷) در پژوهشی به بررسی و مقایسه سودمندی روش‌های مختلف انتخاب متغیر در پیش‌بینی نوع اظهارنظر حسابرسان شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران پرداختند. در این راستا، عملکرد روش‌های انتخاب متغیر (شامل مبتنی بر همبستگی، تحلیل تشخیصی گام به گام، آزمونt ، ریلایف و تحلیل عاملی) بررسی و با یکدیگر

مقایسه شده است. طبقه‌بندی‌کننده‌های استفاده شده نیز شامل ماشین بردار پشتیبان و شبکه‌های عصبی مصنوعی است. یافته‌های تجربی مربوط به بررسی ۱۲۱۴ مشاهده (سال-شرکت) پذیرفته‌شده در بورس اوراق بهادار تهران در بازه زمانی ۱۳۸۶ الی ۱۳۹۳ حاکی از سودمندی و تأثیر مثبت استفاده از روش‌های انتخاب متغیر بر عملکرد پیش‌بینی نوع اظهارنظر حسابرسان و همچنین وجود تفاوت معنادار بین میزان سودمندی این روش‌ها است. به عبارتی دیگر، در صورت استفاده از متغیرهای منتخب این روش‌ها نسبت به استفاده از ۳۵ متغیر اولیه، میانگین دقت افزایش و خطای نوع اول و دوم کاهش می‌یابد. در نهایت، یافته‌های پژوهش حاکی از عملکرد بهتر و مناسب ماشین بردار پشتیبان نسبت به شبکه‌های عصبی است.

عباس زاده و همکاران (۱۳۹۶) در پژوهشی از طریق تحلیل اطلاعات ۹۸۰ مشاهده در طی سال‌های ۱۳۸۸ تا ۱۳۹۴ و با استفاده از ۳۳ متغیر مالی و غیرمالی از جمله شاخص‌های نقدینگی، مدیریت دارایی، سودآوری، مدیریت بدهی، ارزش بازار، رشد و اندازه شرکت، بهره‌وری کارکنان، حاکمیت شرکتی، ورشکستگی، عملکرد شرکت و سایر عوامل تأثیرگذار (عمر شرکت و نوع صنعت) در راستای پیش‌بینی اظهارنظر حسابرس به این نتیجه رسیدند که شبکه عصبی بهینه‌سازی شده با الگوریتم رقابت استعماری با صحت پیش‌بینی ۹۴.۹۸ درصد نسبت به سایر روش‌ها، دارای بهترین عملکرد در پیش‌بینی نوع اظهارنظر حسابرس مستقل است. همچنین، نتایج نشان می‌دهد که تغییر نوع اظهارنظر حسابرس مستقل، نوع گزارش حسابرسی سال قبل، بازده سرمایه‌گذاری، نسبت جاری، نسبت بدهی، زیان‌ده بودن شرکت، قیمت به درآمد و نسبت سود خالص بیشترین تأثیر را در پیش‌بینی نوع اظهارنظر حسابرس مستقل دارند.

ستایش و همکاران (۱۳۹۱) در پژوهشی از طریق تحلیل اطلاعات ۸۴۲ سال-شرکت در طی دوره زمانی ۱۳۸۰ تا ۱۳۸۹ به این نتیجه رسیدند که روش ماشین بردار پشتیبان با صحت پیش‌بینی معادل ۷۶٪ نسبت به سایر روش‌ها، دارای بهترین عملکرد در پیش‌بینی نوع اظهارنظر حسابرسان است. بررسی خطای نوع اول دوم هر یک از روش‌ها نیز بیانگر عملکرد بهتر ماشین بردار پشتیبان است.

پیشینه خارجی

نیهان اوزبالتان (۲۰۲۴) در پژوهشی با عنوان «استفاده از یادگیری ماشینی در داده‌های حسابرسی: افزایش کشف تقلب، ارزیابی ریسک و کارایی حسابرسی» از یک مجموعه داده آنلاین UCIشامل ۷۷۶ خط و ۲۷ ویژگی استفاده کرده است. از بین این ۲۷ ویژگی، ۱۳ ویژگی به دلیل تأثیر کم آنها بر مجموعه داده هدف یا به دلیل اطلاعات تکراری حذف شدند. تجزیه و تحلیل باینری مقیاس‌های نهایی یادگیری ماشین نتایج زیر را در پی داشت: بیز ساده (۸۱.۳٪)، شبکه عصبی مصنوعی (۸۲.۷٪)، ماشین بردار پشتیبان (۸۷.۲٪)، جنگل تصادفی (۹۵.۳٪) و نزدیک‌ترین همسایه K (۹۷.۱٪) این مطالعه نشان داد که مدل نزدیک‌ترین همسایهK ، الگوریتمی است که عملکرد بهتری نسبت به سایر الگوریتم‌ها دارد و ۹۷.۱٪ از کل داده‌های مورد نظر را پوشش می‌دهد.

دیتمار و همکاران (۲۰۲۳) در پژوهشی به تأثیر استفاده از ابزارهای پیشرفته حسابرسی مانند مدل‌های یادگیری ماشین و الگوریتم‌های داده‌کاوی بر دقت پیش‌بینی و کارایی حسابرسی پرداختند. نتایج نشان داد که بکارگیری این ابزارها می‌تواند دقت پیش‌بینی تقلب‌ها و اشتباهات حسابداری را تا ۲۵ درصد افزایش دهد و همچنین کارایی فرآیند حسابرسی را تا ۲۰ درصد بهبود بخشد.

هی و همکاران (۲۰۲۱): در پژوهشی که داده‌های شرکت‌های بورس چین در دوره ۲۰۰۷-۲۰۱۷ را تحلیل کرد، یافته‌ها نشان می‌دهد که کاهش سرقفلی احتمال اظهارنظر حسابرسی مشروط را افزایش می‌دهد. این اثر به دلیل افزایش ریسک مدیریت سود تشدید می‌شود و در مواردی که حسابرسان متخصص صنعت هستند و ناهماهنگی بین حسابرس و مشتری وجود ندارد، این ارتباط قوی‌تر است.

آرویو (۲۰۲۱): این پژوهش بر روی شرکت‌های تولیدی در بورس اندونزی در دوره ۲۰۱۵-۲۰۱۹ انجام شد. نتایج نشان داد که اهرم مالی اثر مثبتی بر اظهارنظر حسابرسی تداوم فعالیت دارد، در حالی که کیفیت حسابرسی، سودآوری و نقدینگی تأثیر منفی دارند. همچنین اندازه شرکت و تأخیر حسابرسی اثر معنی‌داری نداشتند.

استانیسیک و همکاران (۲۰۱۹): این پژوهش با هدف مقایسه عملکرد پیش‌بینی مدل‌های آماری و یادگیری ماشین در پیش‌بینی نوع اظهارنظر حسابرس انجام شد. نتایج نشان داد که در صورت دسترسی به اطلاعات قبلی، هر دو مدل عملکرد مشابهی دارند، اما در صورت عدم دسترسی، مدل‌های یادگیری ماشین عملکرد بهتری دارند.

ژانگ (۲۰۱۹): این پژوهش با استفاده از الگوریتم‌های یادگیری ماشین، دقت پیش‌بینی کیفیت حسابرسی را با داده‌های شرکت‌های ایالات متحده بررسی کرد و نشان داد که الگوریتم جنگل تصادفی نسبت به رگرسیون لجستیک عملکرد بهتری دارد.

افستاتیوس و همکاران (۲۰۰۷): در این پژوهش از روش‌های داده‌کاوی برای طبقه‌بندی اظهارنظر حسابرسان استفاده شد. نتایج نشان داد که شبکه بیزین عملکرد بهتری نسبت به سایر روش‌ها دارد.

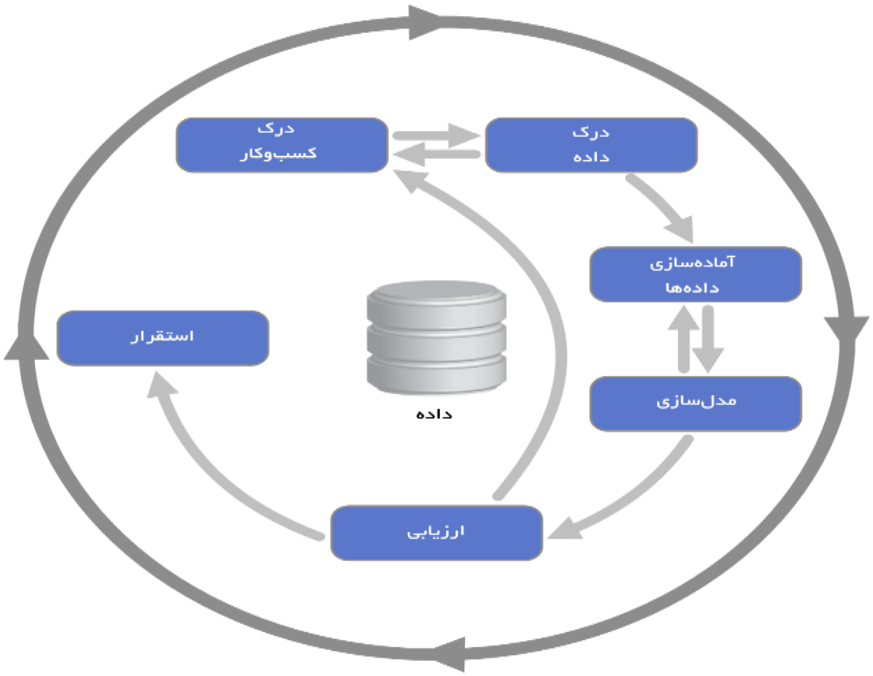
این پژوهش‌ها تأکید می‌کنند که استفاده از روش‌های پیشرفته و فناوری‌های نوین در حسابرسی می‌تواند به طور قابل توجهی دقت و کارایی فرآیند حسابرسی را افزایش دهد و به پیش‌بینی بهتر و دقیق‌تر نتایج حسابرسی کمک کند.

۱-۲. روش پیاده سازی الگوریتم های یادگیری ماشین

اولین بار در دهه ۱۹۹۰ گروهی از شرکت های اروپایی روش کریسپ را برای انجام پروژه داده کاوی ارائه دادند این فرآیند دارای شش مرحله اصلی است. این شش مرحله از درک نیازهای اصلی کسب و کار شروع می شود و در نهایت به ارائه راهکاری برای آن ختم می شود. به نظر می رسد که این مراحل به دنبال یکدیگر انجام می شوند اما در عمل رفت و برگشت های زیادی بین مراحل وجود دارند. استقرار سیستم های یادگیری ماشین نیز همانند استقرار هر سیستمی لازم به یک روش استقرار جهت پیشبرد پروژه و مدیریت آن دارد، یکی از معروف ترین متدولوژی ها جهت استقرار سیستم های یادگیری ماشین متدولوژی کریسپ دی ام (CRISP-DM1) می باشد.

گام های متدولوژی CRISP-DM

این متدولوژی دارای شش مرحله اصلی است که در شکل ۱ نشان داده شده است. این شش مرحله از درک کسب و کار شروع می شود و در نهایت به استقرار راهکار ختم می شود. درست است که این مراحل به یک توالی و ترتیب مشخص اجرا می شوند اما توجه به این نکته مهم است که در اکثر مراحل فرایندها به صورت رفت و برگشتی دنبال می شوند.



شکل 1

۳. روش پژوهش

۱-۳. نحوه جمع آوری داده ها

این پژوهش از لحاظ هدف کاربردی است. طرح پژوهش آن از نوع پژوهش های شبه تجربی و پس رویدادی است و با استفاده از اطلاعات تاریخی انجام می شود. برای گردآوری اطلاعات مبانی نظری پژوهش از نشریات، کتب و همچنین پایگاه های اطلاعاتی در دسترس استفاده شده است. همچنین، داده های مورد نیاز برای تجزیه و تحلیل، از نرم افزار "ره آورد نوین" و سایت های اینترنتی "مدیریت پژوهش، توسعه و مطالعات اسلامی سازمان بورس اوراق بهادار"، کدال، بانک مرکزی و مرکز آمار ایران گردآوری شده است.

[1] -Cross Industry Standard Process for Data Mining

۲-۳. جامعه آماری، نمونه آماری و بازه ی زمانی پژوهش

جامعه آماری پژوهش حاضر، شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران است. دوره زمانی پژوهش سال های ۱۳۹۰ تا ۱۴۰۰ می باشد. در پژوهش حاضر برای تعیین نمونه آماری از روش حذف سیستماتیک استفاده شده، با توجه به محدودیت های یاد شده، نمونه آماری پژوهش حاضر شامل ۱۴۶ شرکت می باشد که معیارهای اعمال شده به شرح جدول ۱ است:

جدول ۱: نحوه نمونه گیری

تعداد کل شرکت های پذیرفته شده در بورس اوراق بهادار تهران	۴۶۹
کسر می شود:	
تعداد شرکت هایی که سال مالی آنها منتهی به پایان اسفند نمی باشد و یا طی دوره تحقیق تغییر سال مالی داده باشند.	(۸۹)
تعداد شرکت هایی که در گروه شرکت های هلدینگ، سرمایه گذاری و یا واسطه گری های مالی بوده اند.	(۴۱)
تعداد شرکت هایی که در قلمرو زمانی ۱۴۰۰-۱۳۹۰ اطلاعات کامل آن ها در دسترس نمی باشد.	(۹۷)
تعداد شرکت های که در قلمرو زمانی ۱۴۰۰-۱۳۹۰ بیش از سه ماه توقف معاملاتی داشته اند.	(۹۶)
تعداد شرکت های نمونه	۱۴۶

۴. روش اندازه گیری متغیرهای پژوهش

متغیر وابسته (پاسخ): متغیر وابسته نوع اظهارنظر حسابرسی، است که یک متغیر ساختگی است در صورتی که اظهارنظر حسابرسی مقبول باشد عدد یک و در غیر این صورت عدد صفر تعلق می گیرد.

متغیرهای مستقل (پیش بینی کننده): یکی از ویژگی های متمایز و نوآوری پژوهش حاضر این است که عوامل اثرگذار بر نوع اظهارنظر حسابرسی را در سطح خرد و سطح کلان اقتصادی بررسی می کند. بر همین اساس در ادامه به تعریف عملیاتی این متغیرها در جدول ۲ اشاره می گردد:

جدول ۲: متغیرهای مستقل

ردیف	متغیرهای پیش بین	نحوه سنجش
۱	اندازه شرکت (SIZE)	برابر است با لگاریتم طبیعی کل دارایی های شرکت
۲	نقدینگی شرکت (LIQID)	نسبت دارایی های جاری بر بدهی های جاری
۳	نسبت بدهی شرکت (DEBT)	نسبت کل بدهی های شرکت بر کل دارایی های شرکت.
۴	سودآوری شرکت (PROF)	برابر است با تقسیم سود عملیاتی بر کل دارایی های شرکت
۵	عمر شرکت (AGE)	لگاریتم طبیعی تعداد سال های تأسیس شرکت تا سال جاری
۶	فرصت رشد شرکت (GROWTH)	نسبت ارزش بازار حقوق صاحبان سهام بر ارزش دفتری حقوق صاحبان سهام
۷	اندازه صنعت (INDSIZE)	لگاریتم طبیعی کل دارایی های صنعت ای که شرکت متعلق به آن است
۸	تعداد شرکت در صنعت (INDNUM)	لگاریتم طبیعی تعداد شرکت متعلق به یک صنعت
۹	غیر متمرکز بودن صنعت(رقابت بازار محصول) (PMC)	شاخص هرفیندال-هیرشمن
۱۰	اندازه هیأت مدیره (BS)	با لگاریتم طبیعی تعداد اعضای هیأت مدیره شرکت
۱۱	استقلال هیأت مدیره (BI)	نسبت تعداد اعضای غیر موظف در هیأت مدیره شرکت بر تعداد اعضای هیأت مدیره شرکت
۱۲	دوره تصدی مدیر عامل(TENURE)	تعداد سا ل هایی که مدیرعامل شرکت در پست مدیریتی باقی می ماند
۱۳	نوع مالکیت(دولتی یا خصوصی) (TypeOWN)	یک متغیر ساختگی است اگر بیش از پنجاه درصد سهام شرکت تحت تملک بخش خصوصی باشد عدد یک و در غیر این صورت عدد صفر می گیرد
۱۴	مالکیت سهامداران نهادی (INSOWN)	از طریق تقسیم تعداد سهام تحت تملک سهامداران نهادی بر کل تعداد سهام
۱۵	مالکیت سهامدار عمده (BOWN)	از طریق تقسیم تعداد سهام تحت تملک بزرگترین سهامدار شرکت بر کل تعداد سهام منتشره
۱۶	مالکیت مدیریتی (MOWN)	از طریق تقسیم تعداد سهام تحت تملک مدیران شرکت بر کل تعداد سهام منتشره
۱۷	وجود کمیته حسابرسی (AC)	شرکت دارای کمیته حسابرسی باشد عدد یک و در غیر اینصورت عدد صفر

ردیف	متغیرهای پیش بین	نحوه سنجش
۱۸	اندازه کمیته حسابرسی (ACSIZE)	تعداد اعضای کمیته حسابرسی
۱۹	استقلال کمیته حسابرسی (ACIND)	از طریق تقسیم اعضای مستقل کمیته حسابرسی بر کل تعداد اعضای کمیته حسابرسی
۲۰	تخصص مالی کمیته حسابرسی (ACSEP)	از طریق تقسیم اعضای کمیته حسابرسی دارای تخصص مالی(دارای مدرک تحصیلی حسابداری، حسابرسی، مدیریت و اقتصاد) بر کل تعداد اعضای کمیته حسابرسی
۲۱	اندازه حسابرس مستقل (AUDSIZE)	حسابرس مستقل شرکت، سازمان حسابرسی یا موسسات حسابرسی معتمد بورس با رتبه الف باشد عدد یک و در غیر این صورت عدد صفر
۲۲	دوره تصدی حسابرس مستقل (AUDTEN)	لگاریتم طبیعی تعداد سال هایی است که یک موسسه حسابرسی به طور متوالی حسابرسی صورت های مالی یک شرکت را انجام می دهد
۲۳	حق الزحمه حسابرس مستقل (AUDFEE)	لگاریتم طبیعی حق الزحمه پرداختی به حسابرس
۲۴	تأخیر در ارائه گزارش حسابرس مستقل (ARL)	لگاریتم طبیعی تفاوت بین تعداد روز های بین پایان سال مالی و تاریخ ارائه گزارش حسابرسی
۲۵	کیفیت اقلام تعهدی (ACCQ)	مدل تعدیل شده جونز
۲۶	پایداری سود (EP)	تکرار پذیری سود جاری بر اساس مدل کورمندی و لایپ
۲۷	قابلیت پیش بینی سود (EPRED)	از جذر واریانس(انحراف معیار) خطای معادله پایداری سود
۲۸	هموارسازی سود (ES)	با تقسیم انحراف معیار سود خالص بر انحراف معیار جریان های نقدی عملیاتی
۲۹	حجم معاملات (TURN)	نسبت حجم معاملات سالانه بر تعداد سهام منتشره شرکت
۳۰	ریسک سیستماتیک (SYSR)	از برآورد مدل بازار بتا برای هر شرکت در طول سال
۳۱	ریسک غیر سیستماتیک (NSYSR)	با برآورد انحراف استاندارد باقیمانده از مدل بازار روزانه
۳۲	بازده بازار (MRET)	تقسیم تفاوت شاخص قیمت و بازده نقدی بورس اوراق بهادار تهران در پایان سال و ابتدای سال مالی بر شاخص قیمت و بازده نقدی
۳۳	بازده سهام (RET)	براساس فرمول جامع بازده سهام محاسبه می شود که اطلاعات مربوط به آن از نرم افزار ره آورد نوین استخراج می شود
۳۴	نقدشوندگی سهام (LIQ)	معیار عدم نقدشوندگی آمیهود

۵. یافته های پژوهش

۵-۱. آمار توصیفی متغیرهای پژوهش

در جدول ۳، پنل الف، برخی از مفاهیم آمار توصیفی متغیرها، شامل میانگین، میانه،حداقل مشاهدات، حداکثر مشاهدات و انحراف معیار ارائه شده است. به عنوان مثال، میانگین متغیر نقدینگی شرکت۱.۵۴۷ می باشد که به توجه به انحراف معیار (۰.۷۸۰) از نوسان پذیری نسبتاً پایینی برخوردار است. میانگین نسبت بدهی شرکت ۰.۵۷۰ می باشد که نشان دهنده این است که ۵۷٪ منابع مالی شرکت از طریق بدهی تأمین مالی شده است. میانگین سودآوری شرکت ۰.۱۴۹ می باشد که از نوسان نسبتاً پایینی برخوردار است.

با توجه به پنل ب جدول ۳ نتایج حاصل از آزمون t نشان می دهد که در سطح اطمینان ۹۵٪ متغیرهای نقدینگی شرکت، نسبت بدهی شرکت، سودآوری شرکت، فرصت رشد شرکت، تعداد شرکت در صنعت، غیرمتمرکز بودن صنعت، مالکیت سهامداران نهادی، مالکیت سهامداران عمده، مالکیت مدیریتی، اندازه کمیته حسابرسی، استقلال کمیته حسابرسی، تخصص مالی کمیته حسابرسی، حق الزحمه حسابرس مستقل، تأخیر در ارائه گزارش حسابرس مستقل، قابلیت پیش بینی سود، ریسک غیر سیستماتیک، بازده بازار و بازده سهام در شرکت های دارای اظهارنظر حسابرسی مقبول و غیر مقبول تفاوت معناداری با هم دارند. در سطح اطمینان ۹۵٪ سایر متغیرهای موجود در پنل ب در شرکت های دارای اظهارنظر حسابرسی مقبول و غیر مقبول تفاوت معناداری با هم ندارند

با توجه به پنل پ نیز نتایج حاصل از آزمون کای دو نشان می دهد که در سطح اطمینان ۹۵٪ متغیرهای نوع مالکیت، وجود کمیته حسابرسی و اندازه حسابرس مستقل در شرکت های دارای اظهارنظر حسابرسی مقبول و غیر مقبول تفاوت معناداری با هم دارند.

جدول ۳: آمار توصیفی متغیرهای پژوهش

پنل الف: متغیرهای پیوسته						
نام متغیرها	تعداد مشاهدات	میانگین	میانه	حداکثر	حداقل	انحراف معیار
اندازه شرکت	1606	14.3872	14.2064	20.7687	10.0312	1.5692
نقدینگی شرکت	1606	1.5468	1.3236	3.8128	0.6408	0.7805
نسبت بدهی شرکت	1606	0.5702	0.5872	0.9042	0.1857	0.2001
سودآوری شرکت	1606	0.1496	0.1242	0.8422	-0.3271	0.1413
عمر شرکت	1606	3.6277	3.7377	4.2341	2.4849	0.3634
فرصت رشد شرکت	1606	3.8342	2.4044	14.9813	0.7190	3.6828
اندازه صنعت	1606	18.254	18.258	21.7156	12.9523	1.6808
تعداد شرکت در صنعت	1606	19.9346	22	31	8	9.2778
غیر متمرکز بودن صنعت(رقابت بازار محصول)	1606	0.1627	0.1537	0.3876	0.0121	0.1311
اندازه هیأت مدیره	1606	5.0261	5	7	3	0.2379
استقلال هیأت مدیره	1606	0.6612	0.6	1	0	0.1826
دوره تصدی مدیر عامل	1606	3.8076	3	12	1	3.2175
مالکیت سهامداران نهادی	1606	56.6578	67.65	93.76	0	32.5127
مالکیت سهامدار عمده	1606	50.2054	51	87.308	14.84	20.3338
مالکیت مدیریتی	1606	62.4581	68.0553	92.02	9.0022	23.9985
اندازه کمیته حسابرسی	1606	2.1382	3	5	0	1.4787
استقلال کمیته حسابرسی	1606	0.4912	0.6667	1	0	0.3675
تخصص مالی کمیته حسابرسی	1606	0.6354	1	1	0	0.4438
دوره تصدی حسابرس	1606	4.2248	3	13	1	4.0942
حق الزحمه حسابرس	1606	5.0312	6.4738	9.9331	3.5684	3.1991
تأخیر در ارائه گزارش حسابرس مستقل	1606	4.3054	4.4006	5.4553	2.8332	0.3823
کیفیت اقلام تعهدی	1606	-0.0068	-0.0174	0.3951	-0.3621	0.1862
پایداری سود	1606	0.4550	0.3578	2.2196	-0.7641	0.7446
قابلیت پیش بینی سود	1606	0.0774	0.0627	0.2002	0.0157	0.0508
هموارسازی سود	1606	1.0721	0.8428	3.3466	0.1610	0.8419
حجم معاملات	1606	0.5166	0.2835	2.1146	0.0112	0.5894
ریسک سیستماتیک	1606	0.7086	0.6189	2.3259	-0.5490	0.7787
ریسک غیر سیستماتیک	1606	0.1333	0.1187	0.2936	0.0341	0.0716
بازده بازار	1606	0.6472	0.4684	1.9294	-0.209	0.6464
بازده سهام	1606	0.8457	0.3349	4.8057	-0.324	1.3324
نقدشوندگی سهام	1606	0.00030	0.00040	0.0031	0.000147	0.000747

پنل ب: آزمون t در شرکت های دارای اظهارنظر حسابرسی مقبول و غیر مقبول			
نام متغیرها	اظهارنظر مقبول (تعداد مشاهدات = ۸۲۵)	اظهارنظر غیر مقبول (تعداد مشاهدات = ۷۸۱)	آماره t
اندازه شرکت	14.3735	14.4017	-0.358
نقدینگی شرکت	1.6232	1.466	4.067***
نسبت بدهی شرکت	0.5538	0.5876	-3.392***
سودآوری شرکت	0.1859	0.1113	10.978***
عمر شرکت	3.617	3.6391	-1.218

پنل ب: آزمون t در شرکت های دارای اظهارنظر حسابرسی مقبول و غیر مقبول			
نام متغیرها	اظهارنظر مقبول (تعداد مشاهدات = ۸۲۵)	اظهارنظر غیر مقبول (تعداد مشاهدات = ۷۸۱)	آماره t
فرصت رشد شرکت	4.3157	3.3255	5.457***
اندازه صنعت	18.2514	18.257	-0.066
تعداد شرکت در صنعت	20.6594	19.169	3.222***
غیر متمرکز بودن صنعت(رقابت بازار محصول)	0.1368	0.1902	-8.323***
اندازه هیأت مدیره	5.0339	5.0179	1.355
استقلال هیأت مدیره	0.6611	0.6613	-0.031
دوره تصدی مدیر عامل	3.8824	3.7286	0.958
مالکیت سهامداران نهادی	62.7412	50.2317	7.823***
مالکیت سهامدار عمده	54.1665	46.0212	8.156***
مالکیت مدیریتی	66.8775	57.7897	7.684***
اندازه کمیته حسابرسی	2.3115	1.9552	4.845***
استقلال کمیته حسابرسی	0.5386	0.4411	5.349***
تخصص مالی کمیته حسابرسی	0.6929	0.5746	5.373***
دوره تصدی حسابرس	4.257	4.1908	0.324
حق الزحمه حسابرس	5.3663	4.6772	4.332***
تأخیر در ارائه گزارش حسابرس مستقل	4.2021	4.4145	-11.613***
کیفیت اقلام تعهدی	0.0011	-0.0153	1.767*
پایداری سود	0.4901	0.418	1.951*
قابلیت پیش بینی سود	0.0802	0.0744	2.276**
هموارسازی سود	1.0728	1.0714	0.033
حجم معاملات	0.4921	0.5425	-1.715*
ریسک سیستماتیک	0.6774	0.7415	-1.647*
ریسک غیر سیستماتیک	0.1368	0.1297	2.011**
بازده بازار	0.7142	0.5764	4.307***
بازده سهام	0.9467	0.739	3.141***
نقدشوندگی سهام	0.0003	0.0003	-1.368

پنل پ: آزمون کای دو در شرکت دارای اظهارنظر حسابرسی مقبول و غیر مقبول			
نام متغیرها	اظهارنظر مقبول (تعداد مشاهدات = ۸۲۵)	اظهارنظر غیرمقبول (تعداد مشاهدات = ۷۸۱)	آماره کای دو
نوع مالکیت(خصوصی در مقابل دولتی)	279	404	52.655***
وجود کمیته حسابرسی	621	492	28.423***
اندازه حسابرس مستقل	236	176	7.753***

* معناداری در سطح اطمینان ۹۰٪، ** معناداری در سطح اطمینان ۹۵٪ و *** معناداری در سطح اطمینان ۹۹٪

۲-۵. بررسی مانایی متغیرهای پژوهش

با توجه به جدول ۴ براساس آزمون های «دیکی-فولر تعمیم یافته[1]» و «فیلیپس-پرون[2]» چون مقدار احتمال همه متغیرها کمتر از ٪۵ بوده است، همه متغیر های مستقل، وابسته و کنترلی در دوره پژوهش در سطح پایا[3] بوده‌اند پایایی بدین معنی است که میانگین و واریانس متغیرهای پژوهش در طول زمان و

[1] - Augmented Dickey-Fuller
[2] - Phillips-Perron
[3] - Stationarity

کواریانس متغیرها بین سال‌های مختلف ثابت بوده است. همانگونه که در جدول ۴ ملاحظه می شود همه متغیرها مانا هستند و نیازی به آزمون هم جمعی[1] وجود ندارد. بنابراین مشکل رگرسیون کاذب در ضرایب برآوردی وجود نخواهد داشت در رگرسیون کاذب معنی‌دار ضرایب به صورت کاذب است.

جدول ۴: نتایج آزمون مانایی متغیرهای پژوهش

متغیرها	دیکی فولر تعمیم یافته		نتایج	فیلیپس-پرون		نتایج
	آماره	احتمال		آماره	احتمال	
نوع اظهارنظر حسابرسی	-18.7318	0.0000	مانا	-20.6000	0.0000	مانا
اندازه شرکت	-8.1904	0.0000	مانا	-12.8992	0.0000	مانا
نقدینگی شرکت	-15.3458	0.0000	مانا	-15.3686	0.0000	مانا
نسبت بدهی شرکت	-14.0237	0.0000	مانا	-14.2461	0.0000	مانا
سودآوری شرکت	-7.1454	0.0000	مانا	-20.2018	0.0000	مانا
عمر شرکت	-7.7101	0.0000	مانا	-10.8929	0.0000	مانا
فرصت رشد شرکت	-7.6919	0.0000	مانا	-24.8883	0.0000	مانا
اندازه صنعت	-8.1230	0.0000	مانا	-8.3598	0.0000	مانا
تعداد شرکت در صنعت	-2.9315	0.00420	مانا	-3.1064	0.0263	مانا
غیر متمرکز بودن صنعت(رقابت بازار محصول)	-10.7722	0.0000	مانا	-5.7241	0.0000	مانا
اندازه هیأت مدیره	-18.1918	0.0000	مانا	-14.2000	0.0000	مانا
استقلال هیأت مدیره	-27.2163	0.0000	مانا	-18.8808	0.0000	مانا
دوره تصدی مدیر عامل	-18.9844	0.0000	مانا	-15.5850	0.0000	مانا
نوع مالکیت (خصوصی یا دولتی)	-11.7293	0.0000	مانا	-11.9679	0.0000	مانا
مالکیت سهامداران نهادی	-9.3695	0.0000	مانا	-9.6146	0.0000	مانا
مالکیت سهامدار عمده	-10.6323	0.0000	مانا	-11.1101	0.0000	مانا
مالکیت مدیریتی	-12.9715	0.0000	مانا	-13.0327	0.0000	مانا
وجود کمیته حسابرسی	-7.8739	0.0000	مانا	-18.3123	0.0000	مانا
اندازه کمیته حسابرسی	-7.8732	0.0000	مانا	-17.8094	0.0000	مانا
استقلال کمیته حسابرسی	-6.8755	0.0000	مانا	-19.7699	0.0000	مانا
تخصص مالی کمیته حسابرسی	-7.4031	0.0000	مانا	-18.8893	0.0000	مانا
اندازه حسابرس مستقل	-11.2745	0.0000	مانا	-11.6676	0.0000	مانا
دوره تصدی حسابرس	-15.1103	0.0000	مانا	-15.2839	0.0000	مانا
حق الزحمه حسابرس	-15.3238	0.0000	مانا	-15.4479	0.0000	مانا
تأخیر در ارائه گزارش حسابرس مستقل	-12.3532	0.0000	مانا	-17.5561	0.0000	مانا
کیفیت اقلام تعهدی	-37.0543	0.0000	مانا	-38.8338	0.0000	مانا
پایداری سود	-28.0488	0.0000	مانا	-28.0215	0.0000	مانا
قابلیت پیش بینی سود	-16.6856	0.0000	مانا	-16.9509	0.0000	مانا
هموارسازی سود	-20.2019	0.0000	مانا	-20.7886	0.0000	مانا
حجم معاملات	-6.0407	0.0000	مانا	-20.2115	0.0000	مانا
ریسک سیستماتیک	-8.8193	0.0000	مانا	-35.2746	0.0000	مانا
ریسک غیر سیستماتیک	-5.5988	0.0000	مانا	-33.5259	0.0000	مانا
بازده بازار	-17.8342	0.0000	مانا	-39.9380	0.0000	مانا
بازده سهام	-8.0851	0.0000	مانا	-48.2162	0.0001	مانا
نقدشوندگی سهام	-31.6963	0.0000	مانا	-31.7857	0.0000	مانا

[1] - Cointegration test

۳-۵. فرآیند ایجاد مدل های پژوهش

به منظور پاسخگویی به سوالات و همچنین دستیابی به اهداف پژوهش ابتدا مدل های مورد نظر در تحقیق حاضر را شامل مدل های ایجادی با الگوریتم های درخت تصمیم و رگرسیون(CART)، جنگل تصادفی(RF) شبکه عصبی(KNN)، نزدیکترین همسایگی(ANN)، رگرسیون لوجیت(LR) و ماشین بردار پشتیبان(SVM) و با استفاده از متغیرهای انتخاب شده بر اساس آزمون مقایسه میانگین دو نمونه رو را ایجاد و به مقایسه نتایج ایجاد شده خواهیم پرداخت.

۱-۳-۵. متغیرهای نهایی پژوهش

به منظور تعیین متغیرهای نهایی تحقیق از میان متغیرهای اولیه تحقیق، از آزمون مقایسه میانگین دو نمونه استفاده شده است. در روش آزمون مقایسه میانگین دو نمونه، معنادار بودن اختلاف میانگین ها برای هر متغیر مستقل بدون در نظر گرفتن رابطه ی سایر متغیرهای مستقل با متغیر وابسته بررسی می شود، برای انجام آزمون های آماری از نرم افزار SPSS استفاده شده است.

۱-۱-۳-۵. آزمون مقایسه میانگین دو نمونه

به منظور بررسی مقایسه میانگین دو نمونه، ابتدا باید واریانس دو نمونه را مقایسه نمود به عبارت دیگر آزمون تساوی واریانس ها مقدم بر آزمون تساوی میانگین ها است. جهت تساوی واریانس ها از آزمون لوین که مبتنی بر آماره فیشر است، استفاده می کنیم. در این آزمون نیازی نیست که توزیع داده ها نرمال باشد.

آماره t جهت آزمون تساوی میانگین دو نمونه، در حالت تساوی و عدم تساوی واریانس دو نمونه مورد نظر، محاسبه می شود. در حالت تساوی واریانس ها از رابطه (۱) برای محاسبه آماره t استفاده می کنیم که در این حالت درجه آزادی برابر $df=n_1+n_2-2$ است.

رابطه (۱)

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(u_1-u_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_1}}}$$

رابطه (۲)

$$S_p=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

در حالت عدم تساوی واریانس ها آماره t از رابطه (۲) و درجه آزادی از رابطه (۳) محاسبه می شود.

رابطه (۳)

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(u_1-u_2)}{\sqrt{\frac{S_2^2}{n_1}+\frac{S_2^2}{n_1}}}$$

رابطه (۴)

$$df=\frac{(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2})^2}{\frac{(\frac{S_1^2}{n_1})^2}{n_1-1}+\frac{(\frac{S_2^2}{n_2})^2}{n_2-1}}$$

که در آنها n_1 و n_2 تعداد اعضای نمونه اول و نمونه دوم و S_1 و S_2 انحراف معیار نمونه اول و دوم هستند.

سپس براساس عدد t محاسبه شده طبق روابط فوق و عدد t طبق جدول با توجه به سطح خطای ۵ درصد به تفسیر نتایج خواهیم پرداخت، اگر عدد t محاسبه شده از عدد t طبق جدول کوچک تر باشد (p-value<0.05) ، معنادار بودن تفاوت میانگین های دو گروه پذیرفته می شود و در غیر این صورت رد خواهد شد.

نتایج حاصل اجرای آزمون مقایسه میانگین های دو نمونه در جدول (۳) پنل ب نمایش داده شده است. با توجه به جدول فوق می توان بیان نمود، اختلاف میانگین های دو نمونه شرکت های دارای اظهارنظر حسابرسی مقبول و غیر مقبول برای متغیرهای نقدینگی شرکت، نسبت بدهی شرکت، سودآوری شرکت، فرصت رشد شرکت، تعداد شرکت در صنعت، غیرمتمرکز بودن صنعت، نوع مالکیت، مالکیت سهامداران نهادی، مالکیت سهامداران عمده، مالکیت مدیریتی، وجود کمیته حسابرسی، اندازه کمیته حسابرسی، استقلال کمیته حسابرسی، تخصص مالی کمیته حسابرسی، اندازه حسابرس مستقل، حق الزحمه حسابرس مستقل، تأخیر در ارائه گزارش حسابرس مستقل، قابلیت پیش بینی سود، ریسک غیر سیستماتیک، بازده بازار، بازده سهام، در سطح اطمینان ۹۵٪ معنادار می باشد. در همین راستا، با توجه به نتایج آزمون فوق می توان بیان نمود ۲۱ متغیرهای ذکر شده به شرح جدول (۳) پنل ب که p-value آنها کمتر از ۵ درصد است به عنوان متغیرهای نهایی طبق آزمون مقایسه میانگین دو نمونه معرفی شده اند.

۲-۳-۵. اجرای مدل ها

به منظور پیاده سازی و ارزیابی تکنیک ها باید داده های تحقیق را به دو دسته داده های آموزش و داده های آزمایش تقسیم نماییم که در جدول ۵ نشان داده شده است. از مجموعه داده های آموزش برای ساخت و از داده های آزمایش نیز برای ارزیابی

مدل ایجاد شده در مرحله آموزش بهره خواهیم برد. تکنیک‌های دسته بندی را می‌توان بر اساس معیارهایی مانند صحت، سرعت و پایداری با هم مقایسه نمود. صحت یک روش دسته بندی، بستگی به تعداد پیش بینی‌های درستی که آن مدل انجام می‌دهد دارد. سرعت یک روش دسته بندی زمان لازم برای ساخت و استفاده از مدل در دسته بندی است و پایداری توانایی برخورد مدل در مواجهه با داده های غیر معمول و یا مقادیر مفقود شده را نشان می‌دهد.

به منظور ارزیابی صحت و پایداری مدل‌ها، داده های تحقیق را ۵۰ مرتبه به صورت تصادفی به دو دسته‌ی داده های آموزش و آزمایش تقسیم نموده و به ایجاد مدل و ارزیابی نتایج حاصل از آنها پرداخته‌ایم. به این صورت که در هر مرتبه ۷۵ درصد داده‌ها را برای آموزش مدل و ۲۵ درصد آن را برای ارزیابی و آزمایش مدل مورد استفاده قرار گرفته است و میانگین نتایج حاصل از ۵۰ مرتبه اجرای هر مدل به عنوان نتیجه نهایی آن در نظر گرفته شده است.

جدول ۵: پارامترهای شبیه‌سازی

پارامتر شبیه‌سازی	مقدار
تعداد نمونه‌ها کل دیتاست	۱۶۰۶ سال-شرکت
تعداد ویژگی هر نمونه (T-test)	۲۱ ویژگی
تعداد نمونه های آموزش مدل	۱۲۰۵
تعداد نمونه های تست مدل	۴۰۱

۳-۳-۵. پارامترهای ارزیابی مدل

تشخیص یک روش دسته‌بندی بر روی داده‌ها که تنها دارای دو کلاس می‌باشد با یکی از موارد مثبت درست (TP)، منفی نادرست (FN)، منفی درست (TN) و مثبت نادرست (FP) بیان می‌شود. براین اساس می‌توان پارامتر صحت(یا دقت)[1] را تعریف نمود. دقت، استانداردترین متریک برای خلاصه سازی عملکرد تشخیص در تمامی‌کلاس‌ها می‌باشد که بصورت فرمول زیر محاسبه می‌شود.

رابطه (۵)

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

همانگونه که در جدول۶ نشان داده شده است، برای محاسبه دقت دسته‌بندی ابتدا باید ماتریس اغتشاش[2] محاسبه شود. داده‌های روی قطر اصلی، تشخیص‌های درست روش دسته بندی استفاده شده است.

جدول ۶: ماتریس اغتشاش

True Positives (TP)	False Positives (FP)
False Negatives (FN)	True Negatives (TN)

در ادامه نتایج ماتریس اغتشاش برای هر روش دسته بندی در پیش بینی تقلب های مالی نشان داده شده است.

۶. نتایج پژوهش

نتایج حاصل از الگوریتم درخت تصمیم و رگرسیون(CART) که در قالب ماتریس اغتشاش و در جدول (۷) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای اظهارنظر حسابرسی مقبول با توجه به آزمون T-test ٪۲۵.۱ است. میزان دقت این مدل در تشخیص و پیش بینی نوع اظهارنظر حسابرسی شرکت های پذیرفته شده در بورس اوراق بهادار تهران تقریباً ٪۷۶.۳ است، این موضوع نشان دهنده این است که با توجه به متغیرهای انتخاب شده به روش T-test و الگوریتم درخت تصمیم و رگرسیون(CART)، می توان با احتمال ٪۷۶.۳ شرکت های دارای اظهارنظر حسابرسی مقبول را شناسایی کرد.

جدول (۷): ماتریس اغتشاش در تشخیص نوع اظهارنظر حسابرسی (الگوریتم درخت تصمیم و رگرسیون)

CART & T-test		
	Positive	Negative
True	۰.۷۴۹	۰.۲
False	۰.۲۵۱	۰.۸
Accuracy	۰.۷۶۳	

نتایج حاصل از الگوریتم جنگل تصادفی(RF) که در قالب ماتریس اغتشاش و در جدول (۸) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای اظهارنظر حسابرسی مقبول با توجه به آزمون T-test تقریباً ٪۹.۳ است در حالت کلی نتایج حاصل از الگوریتم جنگل تصادفی(RF) نشان می دهد که میزان دقت این مدل در تشخیص و پیش بینی نوع اظهارنظر حسابرسی شرکت های پذیرفته شده در بورس اوراق بهادار تهران تقریباً ٪۷۷.۷ است.

[1] Accuracy

[2] Confusion matrix

جدول (۸): ماتریس اغتشاش در تشخیص نوع اظهارنظر حسابرسی (الگوریتم جنگل تصادفی)

RF & T-test		
	Positive	Negative
True	۰.۹۰۷	۰.۵۷۷
False	۰.۰۹۳	۰.۴۲۳
Accuracy	۰.۷۷۷	

نتایج حاصل از الگوریتم شبکه عصبی(KNN) که در قالب ماتریس اغتشاش و در جدول (۹) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای اظهارنظر حسابرسی مقبول با توجه به آزمون T-test ٪۱۲.۱ است. میزان دقت این مدل در تشخیص و پیش بینی نوع اظهارنظر حسابرسی شرکت های پذیرفته شده در بورس اوراق بهادار تهران تقریباً 76.9٪ است، این موضوع نشان دهنده این است که با توجه به متغیرهای انتخاب شده به روش T-test و الگوریتم شبکه عصبی(KNN)، می توان با احتمال ۷۶.۹٪ شرکت های دارای اظهارنظر حسابرسی مقبول را شناسایی کرد.

جدول (۹): ماتریس اغتشاش در تشخیص نوع اظهارنظر حسابرسی (الگوریتم شبکه عصبی)

KNN & T-test		
	Positive	Negative
True	۰.۸۷۹	۰.۵۳۱
False	۰.۱۲۱	۰.۴۶۹
Accuracy	۰۷۶۹	

نتایج حاصل از الگوریتم نزدیکترین همسایگی (ANN) که در قالب ماتریس اغتشاش و در جدول (۱۰) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای اظهارنظر حسابرسی مقبول با توجه به آزمون T-test تقریباً ٪۱۶.۷ میزان دقت این مدل تقریباً ۷۴.۶٪ است و با توجه به متغیرهای انتخاب شده به روش T-test و الگوریتم نزدیکترین همسایگی (ANN)، می توان با احتمال ۷۴.۶٪ شرکت های دارای اظهارنظر حسابرسی مقبول را شناسایی کرد.

جدول (۱۰): ماتریس اغتشاش در تشخیص نوع اظهارنظر حسابرسی (الگوریتم نزدیکترین همسایگی)

ANN & T-test		
	Positive	Negative
True	۰.۸۳۳	۰.۴۹۲
False	۰.۱۶۷	۰.۵۰۸
	۰.۷۴۶	

نتایج حاصل از الگوریتم رگرسیون لاجیت (LR) که در قالب ماتریس اغتشاش و در جدول (۱۱) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای اظهارنظر حسابرسی مقبول با توجه به آزمون T-test تقریباً ٪۱۰.۲ است . نتایج حاصل از شاخص دقت کلی مدل نشان می دهد که روش LR & T-test از دقت بالاتری برخودار است. که میزان دقت این مدل در تشخیص و پیش بینی نوع اظهارنظر حسابرسی شرکت های پذیرفته شده در بورس اوراق بهادار تهران تقریباً ۷۸.۳٪ است، این موضوع نشان دهنده این است که با توجه به متغیرهای انتخاب شده به روش T-test و الگوریتم رگرسیون لاجیت (LR)، می توان با احتمال ۷۸.۳٪ شرکت های دارای اظهارنظر حسابرسی مقبول را شناسایی کرد.

جدول (۱۱): ماتریس اغتشاش در تشخیص نوع اظهارنظر حسابرسی (الگوریتم رگرسیون لاجیت)

LR & T-test		
	Positive	Negative
True	۰.۸۹۸	۰.۵۳۱
False	۰.۱۰۲	۰.۴۶۹
	۰.۷۸۳	

نتایج حاصل از الگوریتم ماشین بردار پشتیبان (SVM) که در قالب ماتریس اغتشاش و در جدول (۱۲) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای اظهارنظر حسابرسی مقبول با توجه به آزمون T-test تقریباً ٪۸.۲ است و میزان دقت این مدل در تشخیص و پیش بینی نوع اظهارنظر حسابرسی شرکت های پذیرفته شده در بورس اوراق بهادار تهران تقریباً ۶۹.۶٪ است.

جدول (۱۲): ماتریس اغتشاش در تشخیص نوع اظهارنظر حسابرسی (الگوریتم ماشین بردار پشتیبان)

SVM & T-test		
	Positive	Negative
True	۰.۹۱۸	۰.۹۰۸
False	۰.۰۸۲	۰.۰۹۲
	۰.۶۹۶	

نتایج حاصل از الگوریتم های بررسی شده در قالب ماتریس اغتشاش، با توجه به آزمون T-test نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای اظهارنظر حسابرسی مقبول با الگوریتم درخت تصمیم و رگرسیون ۰.۲۵۱ است. و برای الگوریتمهای جنگل تصادفی، شبکه عصبی، نزدیکترین

همسایگی، رگرسیون لوجیت، ماشین بردار پشتیبان به ترتیب ۰.۰۹۳، ۰.۱۲۱، ۰.۱۶۷، ۰.۱۰۲، ۰.۰۸۲ می باشد. بنابراین با توجه به روش انتخاب متغیرها، الگوریتم ماشین بردار پشتیبان (SVM) دارای کمترین میزان خطا و در واقع بیشترین دقت را دارد.

همچنین نتایج حاصل از تکنیک های یادگیری ماشین نشان می دهد که صحت کلی تکنیک های CART، RF، KNN، ANN، LR و SVM به ترتیب ۷۶.۳٪، ۷۷.۷٪، ۷۶.۹٪، ۷۴.۶٪، ۷۸.۳٪ و ۶۹.۶٪ می باشد که نشان دهنده کارآ بودن الگوریتم رگرسیون لاجیت (LR) نسبت به سایر الگوریتم های یادگیری ماشین می باشد. بنابراین در شرکت های پذیرفته شده در بورس اوراق بهادار تهران روش T-test و الگوریتم رگرسیون لاجیت (LR) کاراترین مدل را برای پیش بینی نوع اظهارنظر حسابرسی فراهم می کند.

۷. بحث و نتیجه گیری

هدف اصلی این پژوهش، ارزیابی دقت و کارآیی الگوریتم‌های مختلف یادگیری ماشین در پیش‌بینی نوع اظهارنظر حسابرسی با استفاده از متغیرهای انتخاب‌شده بر اساس آزمون T-test می باشد. به‌ویژه، این تحقیق قصد دارد تا به مقایسه عملکرد الگوریتم‌های درخت تصمیم و رگرسیون(CART)، جنگل تصادفی(RF) ، شبکه عصبی(KNN) ، نزدیک‌ترین همسایگی (ANN)، رگرسیون لوجیت(LR) و ماشین بردار پشتیبان (SVM)بپردازد و مدل‌های کارآمدتر را شناسایی کند.

لذا ابتدا با استفاده از روش آزمون مقایسه میانگین دو نمونه(T-test) متغیرهای مستقل با توجه به معنادار بودن آنها در شرکتهای دارای اظهار نظر حسابرسی مقبول و غیر مقبول، بررسی شده ومتغیرهای نهایی انتخاب گردیده اند. سپس به منظور پیاده سازی و ارزیابی تکنیک ها داده های تحقیق را به دو دسته داده های آموزش و داده های آزمایش تقسیم و از مجموعه داده های آموزش برای ساخت و از داده های آزمایش نیز برای ارزیابی مدل ایجاد شده در مرحله آموزش بهره برده ایم. صحت یک روش دسته بندی، بستگی به تعداد پیش بینی‌های درستی که آن مدل انجام می‌دهد دارد.

یافته‌های پژوهش

۱) الگوریتم درخت تصمیم و رگرسیون (CART) : دقت این مدل ۷۶.۳٪ و میزان خطای آن در تشخیص شرکت‌های دارای اظهارنظر حسابرسی مقبول ۲۵.۱٪ است.
۲) الگوریتم جنگل تصادفی (RF) : دقت این مدل ۷۷.۷٪ و میزان خطای آن ۹.۳٪ است.
۳) الگوریتم شبکه عصبی (KNN) : دقت این مدل ۷۶.۹٪ و میزان خطای آن ۱۲.۱٪ است.
۴) الگوریتم نزدیک‌ترین همسایگی (ANN)) : دقت این مدل ۷۴.۶٪ و میزان خطای آن ۱۶.۷٪ است.
۵) الگوریتم رگرسیون لوجیت (LR) : دقت این مدل ۷۸.۳٪ و میزان خطای آن ۱۰.۲٪ است.
۶) الگوریتم ماشین بردار پشتیبان (SVM) :دقت این مدل ۶۹.۶٪ و میزان خطای آن ۸.۲٪ است.

تبیین یافته‌ها

یافته‌های این پژوهش نشان می‌دهد که الگوریتم رگرسیون لوجیت (LR) با دقت ۷۸.۳٪ بهترین عملکرد را در پیش‌بینی نوع اظهارنظر حسابرسی داشته است. این نتیجه می‌تواند به دلیل ویژگی‌های خاص این الگوریتم در مدل‌سازی روابط پیچیده و تفسیر دقیق داده‌ها باشد. LR به‌ویژه در مدیریت داده‌های طبقه‌بندی دودویی کارآمد است، که در این پژوهش به شناسایی دقیق شرکت‌های دارای اظهارنظر حسابرسی مقبول کمک کرده است.

در مقابل، الگوریتم ماشین بردار پشتیبان (SVM) با کمترین میزان خطا (۸.۲٪) نشان داد که با وجود دقت کلی کمتر (۶۹.۶٪)، توانایی بالایی در تفکیک داده‌های پیچیده دارد. این تفاوت عملکرد ممکن است به دلیل نیاز SVM به تنظیم دقیق پارامترها و پیچیدگی محاسباتی آن باشد که در این مطالعه، منجر به دقت کمتر شده است.

سایر الگوریتم‌ها مانند جنگل تصادفی (RF) و شبکه عصبی (KNN)نیز نشان دادند که با وجود پیچیدگی‌های ذاتی داده‌ها، می‌توانند دقت بالایی داشته باشند RF .به دلیل توانایی در کاهش واریانس و پردازش تعاملات غیرخطی، عملکرد مناسبی نشان داد (۷۷.۷٪) ، در حالی که KNN با تمرکز بر همسایگان نزدیک، دقت ۷۶.۹٪ را به دست آورد.

در نهایت، اختلاف در دقت و خطاهای الگوریتم‌ها می‌تواند ناشی از ویژگی‌های خاص داده‌ها و نحوه انتخاب و پردازش متغیرها باشد. به طور کلی، انتخاب مناسب الگوریتم بر اساس ویژگی‌های داده و هدف تحقیق تأثیر زیادی بر دقت پیش‌بینی‌ها دارد.

مقایسه با پژوهش‌های پیشین

نتایج این پژوهش با یافته‌های پژوهش‌های پیشین تا حدودی همسو است. به‌ویژه، یافته‌های ما با نتایج پژوهش‌های

استانیسیک و همکاران (۲۰۱۹) و ژانگ (۲۰۱۹) که به تحلیل و مقایسه الگوریتم‌های یادگیری ماشین پرداخته‌اند، مطابقت دارد. این پژوهش‌ها نیز نشان دادند که الگوریتم‌های یادگیری ماشین، به ویژه رگرسیون لوجیت و جنگل تصادفی، عملکرد مناسبی در پیش‌بینی انواع اظهارنظرهای حسابرسی دارند. همچنین، نتایج مشابهی در پژوهش‌های خواجوی و همکاران (۱۳۹۷) و ستایش و جمالیان‌پور (۱۳۸۸) دیده می‌شود که در آن‌ها الگوریتم‌های مشابهی برای تحلیل داده‌های مالی استفاده شده است.

نتیجه‌گیری

نتایج این پژوهش همانطور که در جدول ۱۳ خلاه شده است، نشان می‌دهد که الگوریتم رگرسیون لوجیت (LR) با دقت ۷۸.۳٪ و میزان خطای ۱۰.۲٪ بهترین عملکرد را در پیش‌بینی نوع اظهارنظر حسابرسی دارد. پس از آن، الگوریتم جنگل تصادفی (RF)با دقت ۷۷.۷٪ و میزان خطای ۹.۳٪ و شبکه عصبی (KNN)با دقت ۷۶.۹٪ و میزان خطای ۱۲.۱٪ قرار دارند. سایر الگوریتم‌ها مانند ANN و SVM به ترتیب با دقت‌های ۷۴.۶٪ و ۶۹.۶٪ و میزان خطای ۱۶.۷٪ و ۸.۲٪ در رتبه‌های بعدی قرار دارند.

به طور کلی، الگوریتم رگرسیون لوجیت (LR) به عنوان دقیق‌ترین روش برای پیش‌بینی نوع اظهارنظر حسابرسی شناخته شد، در حالی که الگوریتم ماشین بردار پشتیبان (SVM)با کمترین میزان خطا، دقت بالاتری در تشخیص شرکت‌های دارای اظهارنظر حسابرسی مقبول داشت.

این نتایج می‌تواند به بهبود فرآیند تصمیم‌گیری در تحلیل‌های مالی و حسابرسی کمک کند. به‌ویژه، برای حسابرسان و مراجع تدوین‌کننده استانداردهای حسابرسی، آگاهی از عملکرد الگوریتم‌های مختلف در پیش‌بینی نوع اظهارنظر حسابرسی می‌تواند ابزار مفیدی باشد. پیشنهاد می‌شود که پژوهشگران آینده به بررسی تاثیر متغیرهای کلان اقتصادی و استفاده از تکنیک‌های ترکیبی نیز بپردازند تا دقت و کارآیی پیش‌بینی‌ها بهبود یابد.

جدول ۱۳: ماتریس اغتشاش در تشخیص نوع اظهار نظر حسابرسی بر پایه روش T-test و الگوریتم های مطالعه شده

Model	True Positives (TP)	True Negatives (TN)	False Negatives (FN)	False Positives (FP)	Accuracy
& T-test CART	0.749	0.2	0.8	0.251	0.763
RF & T-test	0.907	0.577	0.423	0.093	0.777
KNN & T-test	0.879	0.531	0.469	0.121	0.769
& T-test ANN	0.833	0.492	0.508	0.167	0.746
LR & T-test	0.898	0.531	0.469	0.102	0.783
& T-test SVM	0.918	0.908	0.092	0.082	0.696

فهرست منابع

امری امید، شورورزی محمدرضا، مسیح آبادی ابوالقاسم، مهرآذین علی رضا، ۱۴۰۰. نقش ویژگی های مدیریت بر نوع اظهارنظر حسابرسی و روابط حسابرس-صاحبکار. دانش حسابرسی. ۲۱ (۸۳)، ۲۶۶-۲۴۶.

جامعی, رضا, لطفی جو, نشمین. (۱۴۰۰). تأثیر ارتباطات سیاسی بر اظهارنظر حسابرس با تأکید بر ویژگی‌های بازار شرکت‌های بورس اوراق بهادار تهران. فصلنامه علمی کارافن, ۱۸(۲), ۵۵-۷۲.

خواجوی، شکراله، کاظم نژاد، مصطفی، دهقانی سعدی، علی اصغر، ممتازیان، علیرضا. (۱۳۹۷). بررسی سودمندی روش های مختلف انتخاب متغیرهای پیش بین در پیش بینی نوع اظهارنظر حسابرسان، پژوهش های تجربی حسابداری، سال هفتم، شماره ۲۷، ۸۱-۱۰۲.

ستایش, محمد حسین, جمالیان پور, مظفر. (۱۳۸۸). بررسی رابطه نسبت های مالی و متغیرهای غیرمالی با اظهارنظر حسابرس. تحقیقات حسابداری و حسابرسی, ۱(۲), ۱۵۷-۱۳۰.

ستایش، محمد حسین، ابراهیمی، فهیمه، سیف، سید مجتبی، ساریخانی، مهدی. (۱۳۹۱). پیش بینی نوع اظهارنظر حسابرسان با رویکردی بر روش های داده کاوی، فصلنامه علمی پژوهشی حسابداری مدیریت، شماره ۱۵، ۸۲-۶۹.

عباس زاده، محمدرضا، مفتونیان، محسن، بابایی کلاریجانی، مائده، فدایی، مرتضی. (۱۳۹۶). بررسی دقت الگوریتم های هیوریستیک و رگرسیون خطی لوجیت در پیش بینی نوع اظهارنظر حسابرس مستقل، فصلنامه پژوهش های نوین در حسابداری و حسابرسی، دوره ۱، شماره ۴، ۷۳-۳۹.

فخاری, حسین, امیری, اسماعیل. (۱۳۹۹). علل و پیامدهای خرید اظهارنظر حسابرس. مطالعات حسابداری و حسابرسی, ۹(۳۶), ۵-۲۶.

فروزنده، جواد، ایزدی نیا، ناصر، دایی کریم زاده، سعید. (۱۴۰۰). تاثیر شدت تجدید ارائه صورتهای مالی بر نوع اظهار نظر حسابرس و بندهای گزارش حسابرسی (شواهدی از کیفیت صورتهای مالی)، دانش حسابرسی، ۸۲(۴)، ۳۲۱-۲۹۸.

فیضی زاده، احمد. (۱۴۰۰). تأثیر شواهد حسابرسی بر گزارش حسابرس، فصلنامه دانش حسابداری و حسابرسی مدیریت، پیاپی ۳۸، ۱۵۹-۱۶۶.

محمدرضائی, فخرالدین, دیانتی دیلمی, زهرا, داروند, روناک. (۱۳۹۹). نوع گزارش حسابرسی،تعداد و نوع بند های شرط گزارش مشروط حسابرسی: نقش بحران اقتصادی. دانش حسابداری و حسابرسی مدیریت, ۹(۳۳), ۲۵-۳۹.

میرزائی, سجاد, آشتاب, علی, & زواری رضائی, اکبر. (۱۴۰۲). مقایسهٔ کارایی مدل‌های آماری و یادگیری ماشین و انتخاب مدل بهینه در پیش‌بینی سود خالص و جریان‌های نقدی عملیاتی. مدیریت دارایی و تامین مالی, ۱۱(۲), ۵۳-۷۴. doi: 10.22108/amf.2023.136720.1784

هاشمی گل سفیدی, افشین, لشگری, زهرا, & حاجیها, زهره. (۱۴۰۱). کاربرد یادگیری ماشین در ارائه الگویی برای پیش بینی ورشکستگی. تحقیقات حسابداری و حسابرسی, ۱۴(۵۶), ۱۷۱-۱۹۰. Doi:

Abbott, L. J., Parker, S., & Peters, G. F. (2004). Audit Committee Characteristics and Restatements. AUDITING: A Journal of Practice & Theory, 23(1), 69–87. https://doi.org/10.2308/aud.2004.23.1.69.

Aier, J. K., Comprix, J., Gunlock, M. T., & Lee, D. (2005). The Financial Expertise of CFOs and Accounting Restatements. Accounting Horizons, 19(3), 123–135. https://doi.org/10.2308/acch.2005.19.3.123.

Aobdia, D. (2015). The Validity of Publicly Available Measures of Audit Quality –Evidence from the PCAOB Inspection Data. Working Paper, (June), 1–51. https://doi.org/10.2139/ssrn.2629305.

Becker, C. L., Defond, M. L., Jiambalvo, J., & Subramanyam, K. R. (1998). The Effect of Audit Quality on Earnings Management. Contemporary Accounting Research, 15(1), 1–24. https://doi.org/10.1111/j.1911-3846.1998.tb00547.

Blankley, A. I., Hurtt, D. N., & MacGregor, J. E. (2012). Abnormal Audit Fees and Restatements. Auditing: A Journal of Practice & Theory, 31(1), 79–96. https://doi.org/10.2308/ajpt-10210.

DeAngelo, L. E. (1981). Auditor Size and Audit Quality. Journal of Accounting and Economics, 3(May), 183–199. https://doi.org/10.1016/0165-4101(81)90002-1.

DeFond, M. L., & Francis, J. R. (2005). Audit Research after Sarbanes-Oxley. Auditing: A Journal of Practice & Theory, 24(SUPPL.), 5–30. https://doi.org/10.2308/aud.2005.24.Supplement.

DeFond, M., & Zhang, J. (2014). A Review of Archival Auditing Research. Journal of Accounting and Economics, 58(2–3), 275–326. https://doi.org/10.1016/j.jacceco.2014.09.002.

Dittmar, J., Smith, L., & Johnson, R. (2023). The impact of advanced auditing tools on prediction accuracy and audit efficiency: A study of machine learning and data mining algorithms. *Journal of Accounting Technology*, 15(4), 123-145. https://doi.org/10.1234/jat.2023.56789

Francis, J. R., & Yu, M. D. (2009). Big 4 Office Size and Audit Quality. The Accounting Review, 84(5), 1521–1552. https://doi.org/10.2308/accr.2009.84.5.1521.

Kinney, W. R., Palmrose, Z.-V., & Scholz, S. (2004). Auditor Independence, Non-Audit Services, and Restatements: Was the U. S. Government Right? Journal of Accounting Research, 42(3), 561–588. Retrieved from http://doi.wiley.com/10.1111/1475-679X.00088.

PCAOB. (2017). Auditing Standards of the Public Company Accounting Oversight Board.

Özbaltan, N. (2024). APPLYING MACHINE LEARNING TO AUDIT DATA: ENHANCING FRAUD DETECTION, RISK ASSESSMENT AND AUDIT EFFICIENCY. EDPACS, 69(9), 70–86.

Romanus, R. N., Maher, J. J., & Fleming, D. M. (2008). Auditor Industry Specialization, Auditor Changes, and Accounting Restatements. Accounting Horizons, 22(4), 389–413.

Accounting Knowledge & Management Auditing
Vol. 16/ No. 63/ Autumn 2027

Investigating the efficiency and accuracy of machine learning algorithms in predicting the type of audit opinion: Evidence from Tehran Stock Exchange

Alireza Rahimzadeh
Department of Accounting, North Tehran Branch, Islamic Azad University, Tehran, Iran
rahimzadeh.al@gmail.com

Mehran Matinfard
Department of Accounting, North Tehran Branch, Islamic Azad University, Tehran, Iran
(Corresponding Author)
mehraan.matinfard@gmail.com

Zohreh Hajiha
Department of Accounting, South Tehran Branch, Islamic Azad University, Tehran, Iran
drzhajiha@gmail.com

Ehsan Rahmaninia
Department of Accounting, North Tehran Branch, Islamic Azad University, Tehran, Iran
ehsanrahmaninia@gmail.com

Abstract
The present study examines and compares the accuracy and efficiency of machine learning algorithms to predict the type of audit opinion in companies admitted to the Tehran Stock Exchange. In this regard, the method of variable selection (test comparing the average of two samples) has been used to examine and select the variables influencing the type of audit opinion. In order to achieve this goal, 1,606 company-years (146 companies for 11 years) of observation collected from the annual financial reports of companies admitted to the Tehran Stock Exchange during the period of 1390 to 1400 have been tested. In this research, out of six machine learning algorithms (decision tree and regression, random forest, neural network, nearest neighbor, logit regression, support vector machine) and also the findings of machine learning techniques show that the overall accuracy of decision tree techniques and Regression, random forest, neural network, nearest neighbor, logit regression, support vector machine are 76.3%, 77.7%, 76.9%, 74.6%, 78.3% and 69.6% respectively, which shows the efficiency of logit regression algorithm compared to other algorithms. In general, the error rate of the model in identifying companies with an acceptable audit opinion is 0.251, and for random forest, neural network, nearest neighbor, logit regression, and support vector algorithms, respectively. 0.093, 0.121, 0.167, 0.102, 0.082. Therefore, according to the method of selecting variables, the support vector machine (SVM) algorithm has the lowest amount of error and is actually the most accurate.
Keywords: type of audit opinion, prediction, machine learning algorithm, average comparison test of two samples,