Journal of
Professional Auditing Research
Winter 2026, V.6, No 21 pp 192-216

Modeling sample selection in auditing using data mining

Ghodrat allah Barzegar[1], Seyyed Mohammad Bagher Mir Ashrafi[2], abdolhamid sarvedalir[3]

Received: 2025/08/17
Approved: 2025/11/26

Research Paper

Highlights:

• Data mining is used to improve audit efficiency and effectiveness.

• Samples selected through data mining techniques gained greater favor with auditors.

• The combined model of clustering techniques, random forest decision tree, and bootstrapping sampling significantly improved the quality of the selected samples.

• Data mining can be used in other parts of the audit, including audit risk assessment.

Abstract:

Since the selection of the audit sample has a significant impact on the efficiency and effectiveness of the audit process, this research aims to introduce a model for selecting the audit sample using data mining techniques and tries to provide a solution to increase the effectiveness and efficiency of the audit. In this study, first, through interviews with auditors, their needs and knowledge of data mining techniques were assessed. Then, the financial databases of 35 company-years (including 12 companies for the years 1399 to 1401) for which audit files are available were evaluated through various data mining techniques.

The results of the research show that clustering data mining techniques, random decision trees, and bootstrapping sampling can be used as appropriate models for selecting audit samples. These results show that the selected samples, in addition to covering 65% of the samples selected by other methods, can increase the effectiveness of the audit process by providing more relevant samples. On the other hand, more than 63% of auditors were interested in replacing samples selected through data mining techniques with samples selected through traditional methods.

Key Words: Auding Sample Selection, Data mining, Random Decision tree, Clustering, Bootstrapping sampling

Extended Abstract:

1. INTRODUCTION

One of the most important parts of an audit is sample selection, which has a significant impact on the effectiveness and efficiency of the process. Conventional sampling techniques face challenges that can compromise audit quality (Knatchell et al, 2007).

Machine learning is one of the new tools that can be used to improve audit quality and sampling. Data mining, as one of the areas of machine learning, can be used to improve audit

Journal of
Professional Auditing Research
Winter 2026, V.6, No 21 pp 192-216

operations by using advanced algorithms and statistical techniques

2. MATERIALS AND METHODS

Since auditors cannot practically examine all financial transactions, they use sampling to conduct their examinations and generalize the results to the entire population. Auditors can ensure the effectiveness of audit results by more efficiently selecting samples that are prone to error or fraud (Elder and Allen, 2000). Also, selecting the right sample has a significant impact on saving audit time and cost.

Conventional sampling methods in auditing include random, judgmental, stratified, and systematic sampling. All of these methods have been used by auditors over time, but as data sets become larger and more complex, the limitations of these methods become more apparent. These limitations, including human judgment, potential bias, and the inability to dynamically adapt to new data features, have increased interest in advanced approaches such as data mining (Gupta, 2019).

Data mining algorithms, including clustering and random forest decision trees, have been used in modeling financial data for auditing purposes. These tools help manage large-scale financial data, increasing the accuracy of predicting abnormal behavior. Combining these algorithms with bootstrap sampling techniques to improve sample representativeness in auditing has been emphasized in previous research (Awad and Wathik, 2022).

In the first phase of this research, an attempt was made to obtain knowledge about the areas where data mining can be used in audit sampling through interviews with experienced auditors. In semi-structured interviews, auditors answered questions about sampling methods and their level of familiarity with data mining tools. They stated that the size of financial items, relationships between certain accounts in financial balance sheets, and the timing of financial events, especially recurring events, have a great impact on the selection of audit samples.

After conducting interviews with auditors, financial data related to 35 years of the company were examined and, after going through the data preparation steps, clustering techniques, random forest decision trees, and bootstrapping sampling were used on them.

Data mining of financial data was performed using RapidMiner Studio version 9.1. The intuitive user interface, high processing power, and analytical features of this software are effective in data mining of large sets and analyzing data complexity.

3. RESULTS AND DISCUSSION

In implementing data mining algorithms, financial data was divided into 5 clusters with similar characteristics and relationships between certain accounts were extracted by implementing random forest decision tree algorithms. After that, relevant samples were selected from the classifications using the bootstrap technique.

In the first study, the statistical parameters of the selected sample were compared with the parameters of the statistical population, and there was a good match between the parameters of the sample and the statistical population. In some cases, due to the implementation of relationship discovery models, parameters such as the correlation coefficient between certain codes in the debtor and creditor records were improved in the selected sample compared to the statistical population.

After controlling the selected samples through data mining, the new samples were compared with the samples selected by the auditors through traditional techniques, and the comparative

Journal of
Professional Auditing Research
Winter 2026, V.6, No 21 pp 192-216

findings showed that the new samples, in addition to providing adequate coverage of the traditional samples, were able to provide appropriate suggestions regarding samples that could provide a better level of risk coverage for the audit. The quality of the new audit samples was such that about 63 of the new samples were preferred by the auditors to perform the audit work over the traditional samples.

4. CONCLUSION

The selected samples obtained a suitable level of satisfaction among the auditors through data mining techniques. These samples were selected based on the discovery of certain account interrelationships and the discovery of behavioral patterns of recorded financial events, and were able to provide the auditors with appropriate knowledge to perform the audit operations.

This research, with an innovative approach that has not been used in Iran so far, presented a new perspective on the application of data mining techniques in auditing. Despite limitations such as lack of access to companies' financial databases due to confidentiality, future research in the field of audit risk assessment or complementary sample selection methods can reveal more dimensions of the application of data mining in auditing.

10.22034/JPAR.2025.2069241.1456

1. Accounting department, Economics and administrative faculty. University of Mazandaran, babolsar, iran. (Corresponding Author) ghabarzegar@gmail.com
2. Statistics department, mathematics faculty, mazandaran university, babolsar, iran. b.ashrafi@umz.ac.ir
3. Accounting department, economics and administrative faculty, mazandaran university, babolsar, iran. ah.sarvedalir@gmail.com
http://article.iacpa.ir

فصلنامه‌ی پژوهش‌های حسابرسی حرفه‌ای
سال ششم، شماره بیست و یکم، زمستان ۱۴۰۴، صص ۱۹۲-۲۱۶

مدل انتخاب نمونه حسابرسی با استفاده از تکنیک‌های داده‌کاوی

قدرت اله برزگر[۱]✉، سیدمحمد باقر میراشرفی[۲]، عبدالحمید سرودلیر[۳]

مقاله‌ی پژوهشی

تاریخ دریافت: ۱۴۰۴/۰۵/۲۶
تاریخ پذیرش: ۱۴۰۴/۰۹/۰۵

نکات برجسته

- داده‌کاوی برای بهبود کارایی و اثربخشی حسابرسی مورد استفاده قرار می‌گیرد.
- نمونه‌های انتخاب شده از طریق تکنیک‌های داده‌کاوی، مطلوبیت بیشتری نزد حسابرسان بدست آوردند.
- مدل ترکیبی تکنیک‌های خوشه بندی، درخت تصمیم جنگل تصادفی و نمونه‌گیری بوت استراپینگ، کیفیت نمونه‌های انتخاب شده را به نحو قابل ملاحظه بهبود دادند.
- داده‌کاوی در سایر بخش‌های حسابرسی از جمله ارزیابی ریسک حسابرسی می‌تواند مورد استفاده قرار گیرد.

چکیده:

از آنجائی‌که انتخاب نمونه حسابرسی تأثیر بسزائی در کارایی و اثربخشی فرآیند حسابرسی دارد، این پژوهش با هدف معرفی مدلی جهت انتخاب نمونه حسابرسی با استفاده از تکنیک‌های داده‌کاوی سعی در ارائه راهکاری برای افزایش اثربخشی و کارایی حسابرسی دارد. در این پژوهش ابتدا طی مصاحبه‌هایی با حسابرسان، نیازمندی‌ها و دانش آنان از تکنیک‌های داده‌کاوی مورد ارزیابی قرار گرفت. سپس پایگاه‌های داده‌های مالی ۳۵ شرکت-سال (شامل ۱۲ شرکت برای سال‌های ۱۳۹۹ تا ۱۴۰۱) که پرونده‌های حسابرسی آنان در اختیار است، از طریق تکنیک‌های مختلف داده‌کاوی ارزیابی شد.

نتایج تحقیق نشان می‌دهد که تکنیک‌های داده‌کاوی خوشه بندی، درخت تصمیم جنگل تصادفی و نمونه‌گیری بوت استرپینگ می‌توانند به عنوان مدل مناسبی برای انتخاب نمونه حسابرسی مورد استفاده قرار گیرند. این نتایج نشان می‌دهد نمونه‌های انتخاب شده علاوه بر پوشش ۶۵ درصدی نمونه‌های انتخاب شده به روش‌های دیگر، می‌توانند با ارائه نمونه‌های مرتبط‌تر، اثربخشی فرآیند حسابرسی را افزایش دهند. از طرف دیگر بیش از ۶۳ درصد حسابرسان علاقمند بودند تا نمونه انتخاب شده از طریق تکنیک‌های داده‌کاوی را با نمونه‌های انتخاب شده به روش‌های سنتی جایگزین نمایند.

واژه‌های کلیدی: نمونه‌گیری حسابرسی، داده‌کاوی، درخت تصمیم تصادفی، خوشه بندی، نمونه‌گیری بوت استرپینگ.

10.22034/JPAR.2025.2069241.1456

۱. گروه حسابداری، دانشکده علوم اقتصادی و اداری، دانشگاه مازندران، بابلسر، ایران. (نویسنده مسئول) ghabarzegar@gmail.com
۲. گروه آمار، دانشکده علوم ریاضی، دانشگاه مازندران، بابلسر، ایران. b.ashrafi@umz.ac.ir
۳. گروه حسابداری، دانشکده علوم اقتصادی و اداری، دانشگاه مازندران، بابلسر، ایران. ah.sarvedalir@gmail.com

http://article.iacpa.ir

۱- مقدمه

حسابرسی یک عامل کلیدی در اعتباربخشی به صورت‌های مالی است و اعتماد به داده‌های مالی گزارش شده را برای ذینفعان (سرمایه‌گذاران، نهادهای نظارتی و عموم مردم) تأمین می‌کند. فرآیند حسابرسی مستلزم ارزیابی صورت‌های مالی برای تأیید عدم وجود خطاهای اساسی چه به دلیل اشتباه و چه به دلیل تقلب است. یکی از مهمترین جنبه‌های این فرآیند، انتخاب نمونه از مجموعه داده‌های بزرگ است که حسابرسان از آن برای تعمیم ادعاهای مطرح شده در صورت‌های مالی به کل داده‌های مالی استفاده می‌کنند. با این وجود، تکنیک‌های نمونه‌گیری مرسوم با چالش‌هایی در رابطه با اثربخشی، دقت و بی‌طرفی مواجه هستند که می‌تواند به طور بالقوه کیفیت حسابرسی را به خطر بیندازد (نچل، سالتریو و بالو[۱]، ۲۰۰۷).

از طرف دیگر، آینده حسابرسی همچون سایر حوزه‌های علمی، به طور فزاینده‌ای با فناوری‌های هوش مصنوعی و یادگیری ماشینی در هم تنیده شده است. پیکا و زاستمپوفسکی[۲](۲۰۲۵) استدلال می‌کنند که یادگیری ماشینی، شیوه‌های حسابرسی مستمر را ممکن می‌سازد و بینش‌های بلادرنگ و کیفیت حسابرسی بالاتری را ارائه می‌دهد. این تکامل نه تنها فرآیند انتخاب نمونه را ساده می‌کند، بلکه توانایی حسابرس را در تشخیص مؤثرتر تحریفات مالی و تقلب نیز افزایش می‌دهد. علاوه بر این، شرکت‌هایی مانند وست‌راک با موفقیت هوش مصنوعی مولد را در فرآیندهای حسابرسی داخلی خود گنجانده‌اند که منجر به بهبود بهره‌وری و ثبات شده است (دیلویت[۳]، ۲۰۲۴).

اخیراً با داده‌کاوی، رویکرد جدیدی پیش روی حسابرسان قرار گرفته که فرصتی برای بهبود فرآیند انتخاب نمونه در حسابرسی فراهم می‌کند. داده‌کاوی فرآیند یافتن الگوها، همبستگی‌ها و داده‌های پرت در یک مجموعه داده بزرگ با استفاده از الگوریتم‌های پیشرفته و تکنیک‌های آماری است (هان کامبر و پی[۴]، ۲۰۱۱).

نویسندگان در این پژوهش تلاش کرده‌اند تا الگویی از ترکیب تکنیک‌های داده‌کاوی ارائه کنند که می تواند در انتخاب نمونه‌های مفیدتر و مؤثرتر در فرآیند حسابرس مورد استفاده قرار گیرد. نتایج این پژوهش در بررسی پایگاه‌های داده شرکت‌هایی که مورد حسابرسی قرار می‌گیرند، قابل استفاده است.

۲- مبانی نظری و توسعه فرضیه‌ها

انتخاب نمونه از ابتدا یک اصل اساسی حسابرسی بوده است و حسابرسان را قادر می‌سازد تا در مورد مجموعه داده‌های بزرگ، با کارایی نتیجه‌گیری کنند. انتخاب نمونه یکی از فرآیندهای مهمی است که می‌تواند به طور قابل توجهی بر دقت و قابلیت اطمینان حسابرسی تأثیر بگذارد. با توجه به اینکه حسابرسان نمی‌توانند به طور عملی همه تراکنش‌های مالی را در مجموعه داده‌های بزرگ بررسی کنند، از نمونه‌گیری برای بررسی بخشی از داده‌ها و تعمیم نتایج به کل جامعه استفاده می‌کنند. حسابرسان می‌دانند برای یافتن خطرات و بی‌نظمی‌های بالقوه و به دست آوردن

یک نمای کلی، نمونه انتخاب شده باید تا حد امکان مؤثر باشد (آرنز، الدر و بیزلی[۵]، ۲۰۱۴).

علاوه بر این، در مدیریت ریسک، انتخاب نمونه به عنوان مکانیسمی برای هدایت تلاش‌های حسابرسی در مسیرهای پرخطر عمل می‌کند. حسابرسان می‌توانند انتخاب کارآمدتری از نمونه‌هایی که مستعد خطا یا تقلب هستند، انجام دهند تا از مؤثر بودن نتایج حسابرسی اطمینان حاصل کنند (آلن و الدر[۶]، ۲۰۰۰). همچنین در زمان و هزینه حسابرسی صرفه‌جویی می‌شود و در نتیجه آن را برای حسابرس و شرکت مقرون به صرفه‌تر می‌کند (الدر و همکاران[۷]، ۲۰۱۳).

روش‌های قدیمی مورد استفاده در انتخاب نمونه

روش‌های مرسوم انتخاب نمونه، به طور گسترده در حسابرسی به کار گرفته شده‌اند و همه آن‌ها مزایا و محدودیت‌های مربوط به خود را دارند:

نمونه‌گیری تصادفی: اقلام موجود در مجموعه داده‌ها به صورت تصادفی انتخاب می‌شوند. این فرآیند ساده و بی‌طرفانه ممکن است برخی از حوزه‌های پرخطر را از قلم بیندازد، که احتمالاً منجر به نادیده گرفتن مسائل مهم می‌شود (سانتوسو و همکاران[۸]، ۲۰۲۳). علیرغم اینکه حسابرسان ادعا می‌کردند که رویکرد نمونه‌گیری تصادفی می‌تواند فاقد پوشش ریسک‌های بخشی باشد، به صورت کلاسیک از این روش در حسابرس‌ها بویژه در بخش‌های دولتی استفاده کرده‌اند.

نمونه‌گیری قضاوتی: در این روش، حسابرسان از قضاوت حرفه‌ای خود در مورد آنچه که باید در نمونه گنجانده شود، بر اساس‌اندازه تراکنش، ماهیت حساب، تجربه کاری قبلی و غیره، استفاده می‌کنند. اگرچه این روش امکان هدف قرار دادن حوزه‌های پرخطرتر را فراهم می‌کند، اما می‌تواند باعث ایجاد سوگیری و غفلت از شناسایی ریسک‌های نوظهور شود. برخی محققان این رویکرد را به دلیل وابستگی به قضاوت ذهنی مورد انتقاد قرار داده‌اند و از اتخاذ رویکردهای مبتنی بر داده برای کاهش سوگیری حمایت کرده‌اند.

نمونه‌گیری طبقه‌بندی‌شده: در این رویکرد، جامعه بر اساس ویژگی‌های خاص (مثلاًاندازه تراکنش یا نوع حساب) به زیرگروه‌هایی تقسیم می‌شود و نمونه‌ها از هر طبقه انتخاب می‌شوند. این روش مزیت پوشش گسترده‌تر داده‌ها را دارد، اما پیاده‌سازی آن پیچیده‌تر است و نیاز به دانش در مورد ساختار مجموعه داده‌ها دارد(آرنز، الدر و بیزلی[۹]، ۲۰۱۴). این روش نمونه‌گیری در حسابرسی شرکت‌های بزرگ و جایی که شعب در معرض درجات مختلفی از ریسک مالی قرار دارند، مؤثر بوده است. الدر و آلن[۱۰](۲۰۰۰) اشاره کردند، چنین روشی پوشش کلی حسابرسی را از نظر به‌موقع بودن و دقت گزارش‌دهی افزایش می‌دهد.

نمونه‌گیری سیستماتیک: در این روش، حسابرسان نمونه‌هایی را در فواصل تصادفی از لیست مرتب‌شده انتخاب می‌کنند (مثلاً دهمین تراکنش). اگرچه این روش ساده است، اما وقتی روندهای دوره‌ای با فواصل نمونه‌گیری همزمان می‌شوند، می‌تواند الگوهای مهمی را از دست بدهد (نچل و همکاران[۱۱]، ۲۰۰۷).

تمامی روش‌های سنتی اشاره شده در بالا توسط حسابرسان استفاده شده است اما محدودیت‌های آن‌ها با بزرگتر و پیچیده‌تر شدن مجموعه داده‌ها آشکارتر می‌شود. وابستگی به قضاوت انسانی،

سوگیری بالقوه و عدم توانایی در تطبیق پویا با ویژگی‌های داده‌ها، اثربخشی آن‌ها را محدود می‌کند. این امر منجر به علاقه به رویکردهای پیشرفته‌ای مانند داده‌کاوی شده است که دقت و کارایی بهتری را در شناسایی نمونه‌های مناسب ارائه می‌دهند (گوپتا[۱۲]، ۲۰۱۷).

کاربرد فناوری اطلاعات و داده‌کاوی در حسابرسی

نعمتی و همکاران (۱۴۰۴) در پژوهشی تلاش کردند تا با تدوین الگوی بهینه و کاربری فناوری اطلاعات در حسابرسی با توجه به آزمون‌های محتوا، کنترل و ریسک‌های حسابرسی، تمامی عواملی که می‌توانند با کاربرد فناوری اطلاعات در حسابرسی در ارتباط باشند و از آن تأثیر بپذیرند را طبقه‌بندی کنند. آن‌ها به این نتیجه رسیدند که ریسک حسابرسی، آزمون‌های محتوا، نمونه‌گیری و بودجه زمانی بر الگوی بهینه و کاربردی فناوری اطلاعات در حسابرسی تأثیر گذار هستند.

با آشکارتر شدن محدودیت‌های رویکرد سنتی در انتخاب نمونه‌ها، تکنیک‌های داده‌کاوی به طور فزاینده‌ای به عنوان ابزار پیش رو در این حوزه، در نظر گرفته می‌شوند. استفاده از داده‌کاوی می‌تواند ابزارهای ضروری برای رویکرد جدید در تشکیل نمونه را در حسابرسی فراهم کند. با استفاده از الگوریتم‌های پیشرفته‌تر برای تجزیه و تحلیل مجموعه داده‌های بزرگ، داده‌کاوی می‌تواند الگوها و روابطی را شناسایی کند که به راحتی با روش‌های سنتی قابل تشخیص نیستند (امانی و فدلاله[۱۳]، ۲۰۱۷).

تکنیک‌های خوشه‌بندی، طبقه‌بندی و تشخیص ناهنجاری به حسابرسان کمک می‌کند تا مناطق با ریسک بالاتر را با دقت بیشتری شناسایی کنند و منجر به نمونه‌گیری کارآمدتر و مؤثرتر شوند (کوه و تان[۱۴]، ۲۰۱۱).

داده‌کاوی به حسابرسان این امکان را می‌دهد که نمونه خود را در زمان کوتاهتر انتخاب کرده و فرایند حسابرسی را تسهیل نماید (واسارهلی و برون[۱۵]، ۲۰۱۹). این امر به ویژه در زمانی که حسابرسان نیاز به پردازش حجم زیادی از داده‌ها به موقع و دقیق دارند، بسیار مهم است. پیاده‌سازی داده‌کاوی در فرآیند حسابرسی نه تنها کیفیت حسابرسی را بهبود می‌بخشد، بلکه مؤسسات حسابرسی را قادر می‌سازد تا در دنیای همواره در حال تکامل داده‌محور، رقابت‌پذیری خود را حفظ کنند (کاسکارینو[۱۶]، ۲۰۱۲).

شو و لیو[۱۷](۲۰۲۴) نشان دادند که ادغام طبقه‌بندی‌کننده‌های یادگیری ماشین مانند نایو بیز می‌تواند به طور قابل توجهی سوگیری نمونه‌گیری را کاهش دهد و در عین حال حوزه‌های حسابرسی پرخطر را به طور مؤثر شناسایی کند. چنین رویکردهایی حسابرسان را قادر می‌سازد تا منابع را بر روی نمونه‌هایی متمرکز کنند که واقعاً نمایانگر ویژگی‌های کلیدی جمعیت هستند و کیفیت حسابرسی و کارایی را بهبود می‌بخشند.

تکامل تکنیک‌های داده‌کاوی در حسابرسی

پژوهش‌های پیشین که پیشگام کاربرد خوشه‌بندی و تشخیص ناهنجاری در حسابرسی هستند، نتایج امیدوارکننده‌ای را برای به حداقل رساندن سوگیری نمونه‌گیری و افزایش شناسایی تقلب

با اســـتفاده از تکنیک‌های خوشه‌بندی و تشـــخیص ناهنجاری در موسسات مالی بزرگ به دست آورده‌اند.

الگوریتم‌های خوشـــه‌بندی و درخت‌هـــای تصمیم‌گیری جنگل تصادفی بـــه ابزارهای محبوبی در مدل‌ســـازی داده‌های مالی برای اهداف حسابرسی تبدیل شـــده‌اند. خوشه‌بندی به گروه‌بندی تراکنش‌های مالی به بخش‌های معنادار کمک می‌کند، که انتخاب نمونه‌هایی را که تغییرات ذاتی داده‌ها را در بر می‌گیرند، تسهیل می‌کند (جین[۱۸]، ۲۰۱۰). در همین حال، طبقه‌بندی‌کننده‌های درخـــت تصمیم جنـــگل تصادفی در مدیریت داده‌هـــای مالی با ابعاد بالا، اطمینان مناســـبی در خصوص داده‌هـــا ارائه داده و دقت پیش‌بینی را هنگام تشـــخیص ناهنجاری‌ها یا الگوهای جعلی بهبود می‌بخشند (بریمن[۱۹]، ۲۰۰۱). مطالعات اخیر اثربخشی ترکیب این تکنیک‌ها با نمونه‌گیری بوت‌استرپ را برای اصلاح نمایندگی نمونه در زمینه‌های حسابرسی تأیید کرده‌اند (آواد و واتیک[۲۰]، ۲۰۲۲).

چالش‌ها و رویکردهای آتی کاربرد داده‌کاوی در حسابرسی

بـــا وجود مزایای فـــراوان داده‌کاوی، هنگام اجرای آن با چالش‌های متعددی روبرو می‌شـــویم. همانطور که کاســـکارینو (۲۰۱۲) اشـــاره کرد، این تکنیک‌ها نیاز به تخصص خاصی در داده‌کاوی دارند، به این معنی که حسابرسان برای استفاده از این ابزارها باید دارای مجموعه‌ای از مهارت‌های دوگانه در حســـابداری و علوم داده باشـــند. به همین ترتیب، کوه و تان (۲۰۱۱) به مشکلات فنی مرتبط با ادغام سیســـتم‌های داده‌کاوی در فرآیندهای حسابرسی اشاره می‌کنند که به نیروی کار زیادی نیاز دارند و تطبیق فرآیندهای حسابرســـی با فناوری‌های جدید، به ویژه برای شرکت‌های کوچک‌تر، پرهزینه اســـت. آن‌ها اســـتدلال می‌کنند کـــه علیرغم مزایای قابل توجه اســـتفاده از داده‌کاوی برای بهبود کیفیت حسابرسی، زیرساخت‌های فنی و مهارت‌های لازم برای به‌کارگیری مؤثر این تکنیک‌ها در اکثر خدمات حسابرســـی وجود ندارد و تکمیل فرآیند داده‌کاوی از طریق تحقیقات حسابرســـی نیازمند سرمایه‌گذاری بیشـــتر در فناوری و سرمایه انسانی است. همچنین آریشـــودانا و روماه (۲۰۲۳) بحث می‌کنند که حسابرسان باید بر موانع فنی و سازمانی غلبه کنند تا بتوانند به طور کامل از روش‌های داده‌کاوی استفاده کنند.

رویکردهای اســـتاندارد مانند نمونه‌گیری تصادفی، طبقه‌بندی‌شـــده و قضاوتی به خوبی به این حرفه خدمـــت کرده‌اند، اما محدودیت‌هـــای آن‌ها با افزایش‌اندازه مجموعـــه داده‌ها و پیچیدگی طرح‌های کلاهبرداری، آشـــکار می‌شـــود. برای این منظور، به‌کارگیری تکنیک‌های داده‌کاوی با فراهم کردن امکان تجزیه و تحلیل دقیق‌تر بر روی مجموعه بزرگی از داده‌ها، به عنوان یک راه‌حل بالقوه برای این مشـــکلات اثبات شـــده اســـت. در نهایت، روند حرکت از رویکردهای نمونه‌گیری ســـنتی به رویکردهای مبتنی بر داده (داده‌کاوی)، قدرت پیچیدگی داده‌های مالی در کسب‌وکار مدرن را نشـــان می‌دهد و در نتیجه حسابرسی نیز مســـیر حرکتی از میان رویکردهای مبتنی بر داده‌کاوی را دنبال می‌کند.

۳- روش‌شناسی پژوهش

این مطالعه با رویکردی کیفی، تصویر بهتری از شیوه‌های فعلی انتخاب نمونه در حسابرسی ارائه می‌دهد. طرح ریزی اولیه این مطالعه با ایده ابتدایی نویسندگان شکل و با استفاده از ابزارهای هوش مصنوعی بسط یافت لیکن آنچه در این پژوهش مورد استفاده قرار گرفت، تماما توسط نویسندگان تدوین و نتایج آن به قلم ایشان نگارش شده است.

در مسیر اجرای پژوهش، ما با حسابرسان باتجربه از بخش‌های مختلف با استفاده از یک قالب نیمه‌ساختاریافته مصاحبه کردیم. مصاحبه نیمه ساختار یافته یکی از انواع مصاحبه کیفی است که برخی پرسش‌ها از پیش تعیین شده‌اند و در حین اجرا نیز امکان طرح پرسش جدید وجود دارد. این نوع مصاحبه تعادل مناسبی میان چارچوب هدایت شده و انعطاف پذیری ایجاد می‌کند. ما این نوع مصاحبه را در این پژوهش انتخاب کردیم زیرا پژوهشگر می‌تواند براساس پاسخ‌های مصاحبه شونده به سؤالات از پیش تعیین شده، پرسش‌های تکمیلی یا اکتشافی مطرح کند. مصاحبه نیمه ساختار یافته مسیر سوالات را بنحوی هدایت می‌کند که به غنی سازی مدل مفهوم یاری می‌رساند. همچنین مصاحبه، نزدیکی و فاصله همزمان و مناسب را با فضای ذهنی مشارکت کنندگان ممکن می‌کند (معطوفی و همکاران، ۱۴۰۳)

ماهیت نیمه‌ساختاریافته مصاحبه‌ها امکان بررسی عمیق تجربیات و دیدگاه‌های حسابرسان را فراهم کرد. مباحث کلیدی اشاره شده در این مصاحبه‌ها شامل روش‌های نمونه‌گیری کلاسیک مورد استفاده حسابرسان، شاخص‌های انتخاب نمونه‌ها و ارزشگذاری آن‌ها، میزان دانش حسابرسان از تکنیک‌های داده‌کاوی و ارزشمندی آن‌ها و تفکر و برداشت حسابرسان از تکنیک‌های داده‌کاوی بود. ما در این پژوهش از مصاحبه‌شوندگان در مورد رویکرد ترجیحی‌شان برای انتخاب نمونه و همچنین مشکلاتشان هنگام اجرای داده‌کاوی و نظرشان در مورد داده‌کاوی سوال کردیم تا تأثیر آن بر حسابرسی بررسی شود.

ما با ۱۰ حسابرس مصاحبه کردیم که با ترکیبی از سطوح ارشد و مدیران فنی، حسابرسی‌های مختلفی را انجام می‌دادند. در این مصاحبه‌ها از حسابرسانی استفاده شد که علاوه بر انواع سنتی روش‌های حسابرسی، تجربه کار با کلان‌داده و روش‌های حسابرسی مدرن را نیز داشتند.

مجموعه داده‌های مورد رسیدگی

داده‌های مورد رسیدگی در این پژوهش شامل ۳۵ پایگاه داده (شامل ۱۲ شرکت برای سه سال) می‌باشد. بازه زمانی پایگاه‌های داده از سال ۱۳۹۹ تا سال ۱۴۰۱ می‌باشد. داده‌های موجود در پایگاه‌های مذکور شامل کلیه رویدادها و تراکنش‌های مالی در بخش‌های مختلف شرکت‌های مورد رسیدگی شامل فروش، خرید، حقوق و دستمزد، موجودی کالا و سایر سوابق مالی شرکت‌ها می‌باشد.

داده‌های مورد رسیدگی در این پژوهش براساس استانداردهای داده‌کاوی، مورد ارزیابی اولیه به شرح نمودار زیر قرار گرفته‌اند.

شکل ۱ فرآیند ارزیابی اولیه داده‌ها برای داده‌کاوی

استفاده از تکنیک‌های داده‌کاوی

پس از انجام مصاحبه‌ها و جمع‌آوری داده‌ها، سه تکنیک داده‌کاوی مختلف بر روی داده‌های جمع‌آوری‌شده اعمال می‌شود که شامل خوشه‌بندی، درخت‌ها تصمیم‌گیری تصادفی و انتخاب نمونه با استفاده از روش بوت‌استرپ می‌گردد. منطق پشت انتخاب این تکنیک‌ها، توانایی آن‌ها در برخورد با حجم قابل توجهی از داده‌ها و آشکار کردن روابطی است که ممکن است با استفاده از روش‌های سنتی تجزیه و تحلیل به راحتی آشکار نشوند.

۱- الگوریتم K-MEAN برای خوشه‌بندی

یکی از شاخه‌های یادگیری بدون نظارت، خوشه‌بندی است که در آن، به عنوان یک فرآیند خودکار، نمونه‌ها در کلاس‌هایی گروه‌بندی می‌شوند که اعضای آن‌ها مشابه یکدیگر هستند و خوشه نامیده می‌شوند. بنابراین، یک خوشه مجموعه‌ای از اشیاء مشابه است و اشیاء موجود در خوشه‌های مختلف با یکدیگر متفاوتند.

مانند هر تحلیل آماری برای هر داده ورودی، الگوریتم‌های خوشه‌بندی، بدون اینکه از قبل بدانند آیا آن داده‌ها برای خوشه‌بندی مناسب هستند یا خیرو با انتخاب متغیرها، نتایج را در اختیار کاربر قرار می‌دهند (پاستور[۲۱]، ۲۰۱۰). ذکر این نکته ضروری است که در خوشه بندی برای دستیابی به نتایج معتبر، صحت ترتیب بین داده‌ها اهمیت دارد.

نکته مرتبط با این موضوع این است که نباید از متغیرهای زیادی برای خوشه‌بندی مشاهدات استفاده کرد؛ زیرا انتخاب تعداد زیادی از متغیرها برای خوشه‌بندی، احتمال اینکه برخی از متغیرها تقریباً ویژگی‌های یکسانی رااندازه‌گیری کنند، افزایش می‌دهد.

همچنین عواملی که همبستگی بالایی با یکدیگر دارند، نباید تحت خوشه‌بندی یکسان قرار گیرند؛ زیرا نتایج خوشه‌بندی را نسبت به این عوامل حساس‌تر می‌کنند.

در استفاده از الگوریتم خوشه بندی، دو متغیر شماره سند حسابداری و کدهای معین استفاده شده در آن به عنوان متغیرهای اصلی خوشه‌بندی انتخاب شدند و نتایج پیاده‌سازی الگوریتم خوشه‌بندی به الگوریتم درخت تصمیم‌گیری جنگل تصادفی داده شد. دلیل انتخاب حساب‌های معین به عنوان عامل خوشه بندی این است که انتظار می‌رود با انتخاب نمونه مناسب از حساب‌هایی که سطوح خاصی از آن‌ها بخش اصلی سرفصل‌های گزارش‌های مالی را تشکیل می‌دهند، در رسیدگی‌های حسابرسی، اطمینان معقولی در مورد صحت اعداد آن‌ها حاصل شود.

۲- الگوریتم درخت تصمیم‌گیری به روش جنگل تصادفی

آنچه می‌توانیم به عنوان درخت تصمیم در نظر بگیریم، یک ساختار درختی است که در ریشه

هر درخت تصمیم، از قبل یک شرط آزمایشی داریم که دارای زیرشاخه‌هایی است که نشان‌دهنده تصمیم بر اســـاس هر یک از نتایج آن اســـت. هدف این رویکرد، تجزیـــه پیچیدگی این فرآیند با تقسیم مشاهدات به زیرگروه‌های جداگانه است. این روش نسبت به شرایط توزیع داده‌ها و نسبت به متغیرهای ورودی بی‌تفاوت اســـت. روش پیش‌بینی به دلیل تصمیمات متمایز و ســـاده‌ای که به عنوان درخت تصمیم گرفته می‌شـــود، آسان و ساده است. این رویکرد یک واقعیت بسیار مهم، یعنی ســـرعت الگوریتم‌های یادگیری را برجســـته می‌کند. درختان تصمیـــم یکی از ویژگی‌های کلیدی هستند که فرآیند تصمیم‌گیری پیچیده را به تعدادی تصمیم نسبتاً ساده با قابلیت تفسیر تبدیل می‌کنند.

در یادگیری ماشـــین، الگوریتم جنگل تصادفی یک اســـتراتژی یادگیری درخت قوی است که مجموعـــه‌ای از درخت‌های تصمیم‌گیری را در طول مرحله آموزش تولید می‌کند. جنگل تصادفی یک الگوریتم یادگیری گروهی اســـت که با میانگین‌گیـــری از نتایج چندین درخت تصمیم‌گیری آموزش‌دیـــده روی نمونه‌ها و ویژگی‌های تصادفـــی داده کار می‌کند. یکی از مزایای این الگوریتم، کاهش مشکل بیش‌برازش است که عموماً در درخت‌های تصمیم‌گیری منفرد اتفاق می‌افتد.

از آنجایی که رابطه‌ی یکســـانی بین برخی از سرفصل‌های حساب می‌تواند در اسناد مختلف به طور یکسان رخ دهد، مشکل بیش‌برازش می‌تواند بر ایجاد اصول یادگیری در درخت تصمیم‌گیری تأثیر منفی بگذارد، بنابراین، در این مطالعه بر الگوریتم جنگل تصادفی تأکید شد. سپس الگوریتم جنگل تصادفی اجرا شد و نتایج آن برای نمونه‌گیری از یک نمونه‌ی مناسب به الگوریتم بوت‌استرپ ارسال شد.

۳- الگوریتم بازگشتی برای انتخاب نمونه بر اساس بوت‌استرپ

از آنجا که حجم داده‌های مورد نیاز در حسابرســـی بسیار زیاد است، انتخاب نمونه‌ای که بتواند نماینده خوبی از جامعه آماری باشـــد، از اهمیت بالایی برخوردار است. بنابراین، لازم است رویکرد نمونه‌گیری‌ای اتخاذ شـــود که احتمالاً نمونه نسبتاً کوچکی را از حجم زیادی از جامعه آماری که ویژگی‌های مورد نظر جامعه آماری را دارد، انتخاب کند.

بوت اســـتراپینگ یا خودگردان ســـازی را می‌توان انجام نمونه‌گیری با جایگذاری از یک نمونه اصلی به دفعات زیاد دانست. یعنی ما از یک نمونه ثابت با حجم محدود، به دفعات زیاد نمونه‌گیری مجدد با جایگذاری انجام می‌دهیم تا در نهایت بتوان با استفاده از نتایج کلیه دفعات نمونه‌گیری، به یک توزیع نمونه ای مناسب دست یابیم.

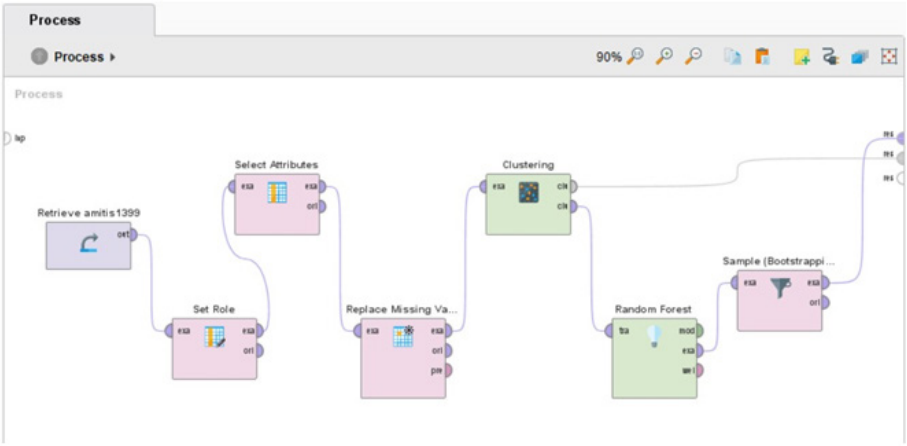
قدرت روش بوت‌اســـترپ در تکنیک‌های علمی و آماری اســـت که می‌تواند برای حل مشکلات مربوط به حجم زیادی از داده‌ها از طریق نمونه‌گیری داده‌های بسیار کم مورد استفاده قرار گیرد. روش بوت‌اســـترپ می‌تواند خطا را با روش نمونه‌گیری مجدد محاســـبه کند و فاصله اطمینان یا انحراف معیار را ارائه دهد. این روش زمانی که تعداد نمونه‌ها کم اســـت و دقت برآوردگرها بسیار مهم است، روشی مناسب و همه‌کاره محسوب می‌شود.

بوت‌اســـترپ خطای استاندارد را تنها بر اســـاس یک نمونه (نمونه اصلی) تخمین نمی‌زند، بلکه

زیرنمونه‌گیری‌های زیادی (در بسیاری از موارد، بیش از ۲۰۰ نمونه) انجام می‌دهد که یک توزیع نمونه آزمایشی تولید می‌کند که برای محاسبه خطای استاندارد استفاده خواهد شد. به گونه‌ای است که توزیع این نمونه آزمایشی به عنوان توزیع نمونه بوت‌استرپ شناخته می‌شود.

تفاوت اصلی که باید به آن اشاره کرد این است که در روش بدون بوت‌استرپ، شرط نرمال بودن توزیع داده‌ها وجود دارد و در استنتاج از طریق نمونه‌گیری مستقل یا بوت‌استرپ چنین شرطی وجود ندارد. یعنی اگر توزیع داده‌های ما نرمال نباشد، از این طریق می‌توان معناداری ضرایب آماری را استنباط و بررسی کرد.

در پژوهش حاضر، خروجی‌های دو الگوریتم خوشه‌بندی و درخت تصمیم بر اساس تحلیل اطلاعات مرتبط در دو الگوریتم قبلی وارد مرحله انتخاب نمونه به روش بوت‌استرپ شده و اسنادی که ارتباط منطقی معقولی با حساب‌های خاص دارند، به عنوان نمونه‌های پیشنهادی انتخاب می‌شوند و با توجه به نمونه کوچکی که انتخاب می‌شود، به عنوان نماینده مناسبی از جامعه آماری انتخاب می‌شوند.



شکل ۲ نحوه چینش الگوریتم‌های داده‌کاوی

ارزیابی روش‌های انتخاب نمونه حسابرس

در مرحله آخر روش‌شناسی، مقایسه‌ای بین تکنیک‌های داده‌کاوی با روش‌های انتخاب نمونه که توسط حسابرسان اعمال می‌شود، ارائه شده است. نویسندگان معتقدند که تجزیه و تکلیل انتخاب نمونه‌ها با استفاه از داده‌کاوی نشان می‌دهد که خروجی‌های حاصل از داده‌کاوی در مقایسه با تکنیک‌های قدیمی از منظر حسابرسان چه میزان مفیدتر و جذابتر هستند.

ابزارها و نرم‌افزارها

روش داده‌کاوی در این مطالعه از طریق RapidMiner Studio (نسخه ۹.۱۰) انجام شد. این نرم‌افزار به دلیل رابط کاربری بصری، قابلیت‌های قدرتمند پردازش داده و ویژگی‌های تحلیلی

پیشـــرفته‌اش که آن را قادر به مدیریت مجموعه داده‌های بـــزرگ و تجزیه و تحلیل‌های پیچیده می‌کند، انتخاب شد.

۴- یافته‌های پژوهش

یافته‌های این پژوهش به ســـه بخش تقسیم می‌شوند. بخش اول آن مرتبط با یافته‌های حاصل از مصاحبه‌ها می‌باشد. در بخش دوم اطلاعات حاصل از جمع آوری و بررسی پایگاه‌های داده بیان می‌گردد. در بخش ســـوم نیز نتایج حاصل از بکارگیری تکنیک‌های داده‌کاوی اشاره شده در بالا، بیان می‌گردد.

۱- آمار توصیفی مصاحبه شوندگان و یافته‌های برتر مصاحبه‌ها

خلاصه نتایج جمعیت شناختی پژوهش در جدول ۱ ارائه شده است. ۸ نفر از مصاحبه‌شوندگان مرد و ۲ نفر آن‌ها زن بودند. از نظر سن، ۷۰ درصد مصاحبه شوندگان با سن زیر ۵۰ سال انتخاب شـــدند تا اطمینان مناســـبی از آشنایی ایشان با ابزارهای نوین بررســـی داده، وجود داشته باشد. همچنین ۸۰ درصد مصاحبه شـــوندگان از افراد دارای تحصیلات کارشناسی ارشد و بالاتر انتخاب شـــدند. رده‌های کاری مصاحبه شوندگان به نحوی انتخاب شد تا تأثیر مستقیم تصمیم ایشان در انتخاب نمونه حسابرسی، قابل بررسی و تحلیل باشد.

جدول ۱ ویژگی‌های جمعیت شناختی پاسخ دهندگان

متغیر	طبقه	فراوانی	درصد	متغیر	طبقه	فراوانی	درصد
جنسیت	مرد	۸	۸۰	رشته تحصیلی	حسابداری و حسابرسی	۸	۸۰
	زن	۲	۲۰		سایر رشته‌ها	۲	۲۰
میزان تحصیلات	کارشناسی	۲	۲۰	رده کاری	حسابرس ارشد	۶	۶۰
	کارشناسی ارشد	۵	۵۰		مدیر فنی	۳	۳۰
	دکتری	۳	۳۰		شریک	۱	۱۰
رده سنی	زیر ۳۹ سال	۴	۴۰	سابقه خدمت	کمتر از ۱۰ سال	۳	۳۰
	بین ۴۰ تا ۴۹ سال	۳	۳۰		بین ۱۰ تا ۲۰ سال	۵	۵۰
	بالای ۵۰ سال	۳	۳۰		بیشتر از ۲۰ سال	۲	۲۰
جمع		۱۰	۱۰۰	جمع		۱۰	۱۰۰

در این پژوهش، مصاحبه‌ها به صورت نیمه ســـاختار یافته طراحی شـــد. ســـوالاتی که در طی

مصاحبه مطرح گردید، طی ۵ روز قبل از تاریخ مصاحبه در اختیار مصاحبه شـــوندگان قرار گرفت تا آن‌ها را بررسی نموده و در خصوص پاسخ آن‌ها تفکر داشته باشند. مدت انجام مصاحبه با توجه به عوامل متعددی از جمله تجربه افراد، رده شـــغلی آن‌ها و ســـطح دانش تخصصی شان متفاوت بوده و تا ۳ ساعت برای اجرای مصاحبه زمان صرف شده است.

پرسش‌های مطرح شده در مصاحبه‌ها در سه دسته کلی تقسیم بندی شدند. بخش اول پرسش‌ها در خصوص روش‌های سنتی نمونه‌گیری در حسابرسی بوده است. در بخش دوم شاخص‌هایی که توســـط حسابرسان برای اجرای نمونه‌گیری مورد اســـتفاده قرار می‌گیرد مورد بررسی قرار گرفت و در بخش ســـوم داده‌کاوی از مصاحبه شوندگان در خصوص میزان آشنایی ایشان با داده‌کاوی و استفاده از تکنیک‌های جدید برای انتخاب نمونه در حسابرسی سوال شد. یافته‌های برتر مصاحبه در جدول ۲ خلاصه شده‌اند.

جدول ۲ مروری بر مصاحبه‌های حسابرسان

موضوع	یافته‌های کلیدی
روش‌های نمونه‌برداری سنتی	نمونه‌گیری تصادفی، نمونه‌گیری طبقه‌بندی شده و نمونه‌گیری قضاوتی، روش‌های سنتی رایج در انتخاب نمونه حسابرسی هستند
شاخص‌های انتخاب نمونه و اولویت‌های آنان	۱- مبلغ و زمان تراکنش ۲- سهم هر سند حسابداری از کل مبالغ هر سرفصل معین ۳- رابطه متقابل حساب‌های معین در طرفین بدهکار و بستانکار اسناد مالی
آگاهی از داده‌کاوی	دانش نسبت به داده‌کاوی کم بود لیکن برای آشنایی با تکنیک‌های جدید علاقمند بودند.
نظرات در مورد داده‌کاوی	۱- دیدگاه در خصوص استفاده از تکنیک‌های جدید نمونه گیری مثبت بوده و آن را مفید می‌دانستند. ۲- در خصوص پیچیدگی استفاده و تحلیل و هزینه‌های بکارگیری این تکنیک‌ها، نگرانی‌هایی وجود داشت.

۲- آمار توصیفی و نتایج جمع آوری و بررسی پایگاه‌های داده

خلاصه نتایج جمع آوری و بررســـی پایگاه‌های داده در جداول ۳ تا ۵ ارائه شده است. همانگونه که پیشـــتر اشاره شـــد، در این تحقیق از ۳۵ پایگاه داده برای ۱۲ شرکت طی سال‌های ۱۳۹۹ تا ۱۴۰۱ اســـتفاده شـــد. دلایل اصلی انتخاب پایگاه‌های داده برای این دوره‌های مالی به شرح زیر است:

اول: از زمان حسابرســـی آن‌ها فاصله زمانی زیادی نگذشته باشد تا با رجوع به پرونده‌های آنان و پرســـش از حسابرســـان مربوطه، امکان دریافت قضاوت‌های حرفه ای حسابرسان مربوطه وجود داشته باشد.

دوم: گزارش حسابرسی مربوط به دوره مذکور صادر شده و پرونده مورد بررسی کیفیت مؤسسه

و ارکان نظارتی قرار گرفته باشـــد تا از صحت پرونده‌های حسابرسی و کیفیت اطلاعات آن بتوان اطمینان معقول حاصل نمود.

در انتخاب پایگاه‌های داده تلاش شـــد که شـــرکت‌هایی مورد بررسی قرار گیرند که در صنایع مختلف فعالیت دارند و فرآیندهای تجاری آن‌ها دارای وابســـتگی و همبســـتگی با شرایطی خاص اقتصادی از جمله دوره‌های رکود و صعود، وابســـتگی‌های تجاری ارزی و ســـایر شـــرایط مؤثر بر رویدادهای مالی آن‌ها نباشـــد. همچنین شرکت‌هایی در ارزیابی پایگاه‌های داده انتخاب شدند که بخش عمده تراکنش‌های مالی از جمله خرید، فروش، چرخه‌های حساب‌های دریافتنی و پرداختنی، تولید، موجودی‌های کالا، دارایی‌های ثابت و تســـهیلات را داشته باشند. کلیه تراکنش‌های مالی و مبالغ درج شده در جداول زیر به واحد «میلیون ریال» می‌باشد.

جدول ۳ آمار توصیفی پایگاه‌های داده شرکت‌ها

سال	تعداد پایگاه داده	مجموع تعداد تراکنش‌ها	مجموع مبلغ تراکنش‌ها
۱۳۹۹	۱۱	۶۵۰,۴۲۰	۳۶۳,۸۴۳,۸۰۷
۱۴۰۰	۱۲	۷۰۸,۰۲۶	۴۲۸,۷۱۹,۸۸۸
۱۴۰۱	۱۲	۴۹۳,۸۳۵	۶۴۴,۰۸۷,۴۹۱
جمع	۳۵	۱,۸۵۲,۲۸۱	۱,۴۳۶,۶۵۱,۱۸۶

جدول ۴ آمار توصیفی تعداد تراکنش‌ها در پایگاه‌های داده شرکت‌ها

سال	میانگین	کمینه	بیشینه	انحراف معیار
۱۳۹۹	۵۹,۱۲۹	۴,۸۱۵	۲۴۷,۹۵۴	۶۶,۲۹۴
۱۴۰۰	۵۹,۰۰۲	۵,۷۲۹	۲۴۱,۸۶۷	۶۱,۵۰۳
۱۴۰۱	۴۱,۱۵۳	۶,۰۶۱	۹۱,۳۵۱	۳۱,۰۳۷

جدول ۵ آمار توصیفی مجموع مبلغ تراکنش‌ها در پایگاه‌های داده شرکت‌ها

سال	میانگین	کمینه	بیشینه	انحراف معیار
۱۳۹۹	۳۳,۰۷۶,۷۱۰	۶۰۸,۳۰۹	۹۳,۹۶۴,۳۴۴	۳۵,۶۷۹,۷۳۷
۱۴۰۰	۳۵,۷۲۶,۶۵۷	۱,۱۲۸,۴۶۸	۱۱۲,۳۳۰,۴۳۵	۳۸,۷۸۹,۰۲۳
۱۴۰۱	۵۳,۶۷۳,۹۵۸	۱,۹۴۸,۰۰۱	۱۶۲,۵۶۷,۳۸۲	۵۵,۰۲۹,۲۱۶

۳- نتایج تکنیک‌های داده‌کاوی

بـــرای آزمایش توانایی روش‌هـــای داده‌کاوی در ارزیابی معیارهای انتخـــاب نمونه از مجموعه داده‌های موجود، از نرم‌افزار Rapidminer اســـتفاده شد . تکنیک‌های مختلف داده‌کاوی هر یک از شاخص‌های مورد نظر حسابرس را به شرح زیر ارزیابی کردند.

۳-۱- خوشه بندی جهت ارزیابی از نظر زمان، مقدار و دفتر کل حساب:

ما خوشه‌بندی را برای ارزیابی رابطه (بر اساس مقدار و زمان هر رکورد حسابداری) بین حساب‌ها بر اساس نظرات جمع‌آوری‌شده از حسابرسان انجام دادیم. داده‌هایی که در یک خوشه با یک ویژگی مشترک قرار می‌گیرند، می‌توانند به عنوان نماینده خوشه استفاده شوند و تکنیک‌های خوشه‌بندی به ما امکان می‌دهند اطلاعات را با استفاده از یک ویژگی مشترک خوشه‌بندی کنیم. برای این مجموعه داده تکمیل‌شده، ما با ۵، ۷ و ۱۰ خوشه آزمایش کردیم و نتایج را مقایسه کردیم. از آنجایی که افزایش تعداد خوشه‌ها تأثیر معنی‌داری بر تعداد نمونه‌ها در هر خوشه نداشت، حداقل ۵ خوشه تأیید شدند.

Cluster Model

```
Cluster 0: 1524 items
Cluster 1: 1690 items
Cluster 2: 1458 items
Cluster 3: 1665 items
Cluster 4: 1585 items
Total number of items: 7922
```

شکل ۳ نمونه خروجی خوشه‌بندی برای مجموعه داده‌های یک ساله شرکت

۳-۲- درخت تصمیم ارزیابی‌شده‌ی هر خوشه:

پس از یافتن خوشه‌های داده با استفاده از روش‌های خوشه‌بندی، ویژگی‌های مربوطه که توسط درخت تصمیم الگوریتم جنگل تصادفی تعریف شده‌اند، ارزیابی شدند. در شکل شماره ۳ بخشی از یکی از درخت‌های تصمیم ایجاد شده در یکی از خوشه‌ها آورده شده است.

RegressionTree

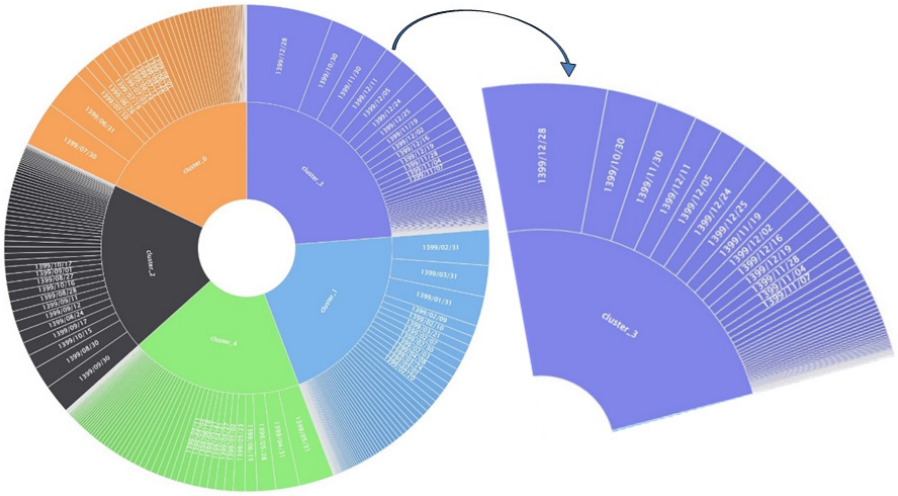
```
4994< بدهکار
|   125423100< بدهکار
|   |   1758< شماره سند
|   |   |   7361.000: 1817.500< شماره سند  {count=5}
|   |   |   1817.500≥ شماره سند
|   |   |   |   389376500< بدهکار
|   |   |   |   |   1760< شماره سند
|   |   |   |   |   |   429376500< بدهکار
|   |   |   |   |   |   |   1795.500< شماره سند
|   |   |   |   |   |   |   |   2534.667: 1799.500< شماره سند  {count=3}
|   |   |   |   |   |   |   |   8102.000: 1799.500≥ شماره سند  {count=2}
|   |   |   |   |   |   |   1795.500≥ شماره سند
|   |   |   |   |   |   |   |   1867.667: 611914738< بدهکار  {count=3}
|   |   |   |   |   |   |   |   3251.000: 611914738≥ بدهکار  {count=2}
|   |   |   |   |   |   |   8403.000: 4293765002 بدهکار  {count=1}
|   |   |   |   |   |   8999.000: 17602 شماره سند  {count=2}
|   |   |   |   |   389376500≥ بدهکار
|   |   |   |   |   |   269621875< بدهکار
|   |   |   |   |   |   |   3201.000: 299994000< بدهکار  {count=3}
|   |   |   |   |   |   |   1040.000: 299994000≥ بدهکار  {count=1}
|   |   |   |   |   |   3201.000: 269621875≥ بدهکار  {count=7}
```

شکل ۴ تفکیک اسناد هر خوشه با استفاده از مدل درخت تصمیم

۳-۳- انتخاب نمونه با استفاده از الگوریتم بوت‌استرپ:

میزان نمونه ای که باید انتخاب شود، ارتباط مستقیمی با ریسک قابل پذیرش حسابرسی دارد. لذا جهت انتخاب نمونه، ابتدا میزان ریسک قابل پذیرش از طریق کاربرگ‌های حسابرسی تعیین و متناسب با آن، حجم نمونه انتخابی مشخص می‌گردد.

پس از تعیین حجم نمونه انتخابی، الگوریتم انتخاب نمونه بوت‌استرپ اعمال شد که نمایش کمی آن را می‌توان در تصویر زیر مشاهده کرد.



شکل ۵ تاریخ و خوشه‌بندی بر اساس انتخاب نمونه

در شکل شماره ۵، ترکیب نمونه‌های انتخاب شده براساس خوشه و تاریخ نمایش داده شده است. هر رنگ نشان دهنده یک خوشه بوده که نام خوشه در برش داخلی رنگ دیده می‌شود. در برش بیرونی رنگ‌ها نیز تاریخ‌هایی که شامل نمونه انتخاب شده هستند، نمایش داده شده است. عرض هر مقطع تاریخی نشان دهنده حجم نمونه انتخاب شده در آن تاریخ است. در واقع هر چه مقطع برشی هر تاریخ بزرگتر باشد، میزان نمونه انتخاب شده در آن تاریخ بیشتر می‌باشد.

همانگونه که در تصویر دیده می‌شود، تعداد نمونه انتخاب شده در روزهای پایانی هر ماه بیشتر است. براساس این تصویر استدلال می‌شود که احتمالا حجم اسناد صادر شده در روزهای پایانی ماه نسبت به روزهای ابتدایی ماه بیشتر است. از طرف دیگر می‌توان استدلال کرد که حجم روابط بین حساب‌های معین یا رخدادهای خارج از الگوهای کشف شده در بازه‌های زمانی پایان ماه، نسبت به سایر روزها بیشتر است. همچنین براساس بررسی دسته بندی درج شده در تصویر استدلال می‌شود حجم اسناد در ماه‌های پایانی سال بیشتر از زمان‌های دیگر در سال بوده که باعث شده است نمونه‌های انتخابی در آن زمان نیز از روزهای دیگر سال بیشتر باشد.

شکل ۶ دفتر کل حساب و نمودار انتخاب نمونه مبتنی بر خوشه

در شـــکل شماره ۶، نحوه توزیع نمونه‌های انتخاب شـــده براساس خوشه‌ها (رنگ‌های ۵ گانه)، کد حساب‌های معین و شماره اسناد می‌باشد. در این شکل محور افقی شامل شماره اسناد، محور عمودی شـــامل کدهای معین و نقاط داخل نمودار، نمونه‌های انتخاب شـــده هستند که رنگ هر نقطه نشان‌دهنده خوشه‌ای است که اسناد در آن طبقه‌بندی شده‌اند.

براســـاس این شکل مشاهده می‌شود که عمده اسناد انتخاب شده مربوط به کدهای وجوه نقد، حســـاب‌های دریافتنی و پرداختنی و هزینه‌ها می‌باشـــند. همچنین بخشی با تراکم کمتر مربوط به اســـناد درآمدی می‌باشـــد. از اطلاعات دریافتی از شکل استدلال می‌شود که عمده روابط میان حســـاب‌های معین متقابل، مرتبط با سرفصل‌های دفتر کل اشاره شده در بالا است و الگوی درج و اســـتفاده از حســـاب‌های معین را به نمایش می‌گذارد و می‌توان برآورد نمود که با رسیدگی به سرفصل‌های مذکور، بخش مهمی از ریسک حسابرسی را پوشش داد.

نتایج داده‌کاوی در مقایسه با نتایج انتخاب نمونه سنتی

پس از اجرای تکنیک‌های داده‌کاوی و انتخاب نمونه جدید با استفاده از این تکنیک‌ها، تحلیل‌ها و مقایسه‌هایی به شرح زیر صورت پذیرفت:

۱- مقایسه ویژگی‌های نمونه انتخاب شده از طریق داده‌کاوی با جامعه آماری مربوطه:

پس از تعیین نمونه انتخاب شـــده از طریق پیاده ســـازی تکنیک‌هـــای داده‌کاوی، ویژگی‌های آمـــاری مربوط به این نمونه‌ها با ویژگی‌های جامعه آماری مربوط به هر نمونه بررســـی گردید که نتایج حاصل از آن به شرح جدول ۶ می‌باشد. میانگین مربوط به مبالغ تراکنش‌های مالی و ضریب همبســـتگی مربوط به شماره حســـاب معین تراکنش‌های مالی در دو ســـمت بدهکار و بستانکار می‌باشد.

براســـاس اطلاعات خلاصه شـــده در جدول ۶ نمونه‌های انتخاب شـــده براســـاس تکنیک‌های

داده‌کاوی، دارای ویژگی‌های آماری مشـــابه به کل جامعه اطلاعات مالی ثبت شده در پایگاه‌های شرکت‌ها هســـتند. در برخی موارد از جمله رشد ضریب همبستگی حساب‌های معین در طرفین ثبـــت یک رویداد مالی، نشـــان از شناســـایی برخی روابـــط موجود در ثبت‌هـــای مالی از طریق تکنیک‌های داده‌کاوی شـــده که این مســـأله می‌تواند به بهبود نتایج ارزیابی نمونه استخراج شده از طریق داده‌کاوی، منجر شـــود. کلیه تراکنش‌های مالی و مبالغ درج شـــده در جدول ۶ به واحد «میلیون ریال» می‌باشد.

۲- مقایســـه ویژگی‌های نمونه انتخاب شده از طریق داده‌کاوی با نمونه انتخاب شده طی روال عادی حسابرسی:

در جدول شـــماره ۷ نتیجه حاصل از نمونه‌گیری با اســـتفاده از تکنیک‌های داده‌کاوی با نتایج حاصل از روش‌های ســـنتی مقایســـه شـــد. پس از انتخاب نمونه با تکنیک‌های داده‌کاوی، نمونه مورد نظر از هر سال-شـــرکت به حسابرسان داده شـــد تا تجزیه و تحلیل مقایسه‌ای با نمونه‌های انتخاب‌شـــده با روش‌های ســـنتی حسابرســـی، در مورد نرخ پذیرش انتخاب نمونه انجام دهند. همچنین از آن‌ها خواسته شد تا میزان تمایل خود را برای جایگزینی نمونه‌های جدید با نمونه‌های انتخاب‌شده قبلی بیان کنند.

جدول ۶ مقایسه ویژگی‌های نمونه جدید با جامعه مورد رسیدگی

شماره پایگاه	حجم تراکنش‌ها		میانگین		ضریب همبستگی	
	جامعه	نمونه	جامعه	نمونه	جامعه	نمونه
۱	۶۱,۹۹۰	۱۵,۴۹۸	۱,۱۲۲	۱,۱۸۹	۶۹	۷۲
۲	۷۰,۲۰۵	۲۴,۵۷۲	۱,۳۲۶	۱,۳۶۳	۷۲	۷۷
۳	۴,۸۱۵	۱,۶۸۵	۱,۲۵۳	۱,۳۲۷	۸۳	۹۴
۴	۵,۷۲۹	۱,۴۳۲	۱۷	۱۷	۶۵	۷۳
۵	۷,۷۸۶	۳,۵۰۴	۲,۲۸۲	۲,۴۱۵	۷۴	۹۵
۶	۶,۰۶۱	۲,۴۲۴	۸۳۵	۸۵۷	۵۵	۸۶
۷	۱۶,۰۰۳	۵,۶۰۱	۹۹۹	۱,۰۵۶	۸۵	۸۶
۸	۷,۹۲۲	۲,۳۷۷	۳۱۷	۳۲۵	۵۳	۹۴
۹	۱۱,۶۵۹	۲,۹۱۵	۴۷۶	۵۰۳	۷۶	۹۳
۱۰	۲۴۷,۹۵۴	۷۴,۳۸۶	۱۲۵	۱۲۸	۸۸	۹۰
۱۱	۲۴۱,۸۶۷	۸۴,۶۵۳	۶۶	۷۰	۶۷	۷۱
۱۲	۲۱,۵۸۴	۹,۷۱۳	۱۰۵	۱۰۷	۹۴	۹۵
۱۳	۳۸,۹۴۶	۷,۷۸۹	۲,۴۱۳	۲,۵۴۴	۶۷	۸۱
۱۴	۳۶,۵۴۰	۹,۱۳۵	۷۵۸	۷۷۵	۷۲	۷۸
۱۵	۴۲,۰۵۰	۱۲,۶۱۵	۱۲۷	۱۳۴	۵۸	۶۹
۱۶	۷۵,۲۴۶	۲۶,۳۳۶	۹۱۴	۹۳۴	۷۹	۸۱

شماره پایگاه	حجم تراکنش‌ها		میانگین		ضریب همبستگی	
	جامعه	نمونه	جامعه	نمونه	جامعه	نمونه
۱۷	۷۰,۵۴۷	۲۱,۱۶۴	۱,۲۵۶	۱,۳۲۲	۵۶	۸۱
۱۸	۶۸,۵۷۵	۲۰,۵۷۳	۱,۸۴۹	۱,۸۸۷	۵۵	۶۱
۱۹	۱۲,۳۰۵	۴,۹۲۲	۱۴۰	۱۴۷	۵۲	۷۵
۲۰	۱۱,۴۴۵	۴,۰۰۶	۱۵۱	۱۵۴	۸۱	۸۶
۲۱	۱۱,۲۶۳	۲,۸۱۶	۱۷۳	۱۸۲	۸۴	۹۴
۲۲	۸۰,۸۸۶	۲۰,۲۲۲	۱,۰۴۳	۱,۰۶۳	۹۴	۹۵
۲۳	۸۵,۸۵۶	۳۰,۰۵۰	۹۵۸	۱,۰۰۵	۵۱	۹۲
۲۴	۹۱,۳۵۱	۳۱,۹۷۳	۱,۷۷۹	۱,۸۱۱	۸۹	۹۱
۲۵	۶۳,۱۵۹	۲۵,۲۶۴	۱,۰۳۶	۱,۰۸۶	۶۳	۸۴
۲۶	۶۴,۷۷۱	۲۹,۱۴۷	۱۷,۳۰۴	۱۷,۶۰۱	۸۳	۸۶
۲۷	۶۹,۵۷۶	۱۷,۳۹۴	۱,۷۸۶	۱,۸۷۱	۷۲	۸۲
۲۸	۶,۴۸۱	۱,۹۴۴	۱۴۱	۱۴۳	۸۸	۸۹
۲۹	۹,۳۶۳	۳,۲۷۷	۲۷۳	۲۸۶	۸۶	۹۲
۳۰	۱۰,۳۶۸	۲,۰۷۴	۴۴۹	۴۵۶	۵۶	۹۲
۳۱	۷۸,۴۹۲	۱۹,۶۲۳	۱۰۷	۱۱۲	۶۱	۷۲
۳۲	۷۲,۸۱۶	۱۴,۵۶۳	۱۵۸	۱۶۰	۷۵	۸۲
۳۳	۷۹,۰۱۳	۲۷,۶۵۵	۱۹۸	۲۰۷	۸۹	۹۵
۳۴	۳۴,۲۱۴	۱۳,۶۸۶	۱۸	۱۸	۶۷	۹۱
۳۵	۳۵,۴۴۳	۱۴,۱۷۷	۳۲	۳۳	۵۵	۶۳

پس از بررسی و تطبیق نمونه‌های ایجاد شده از طریق بکارگیری تکنیک‌های داده‌کاوی، نظر حسابرسان در خصوص اثرگذاری نمونه‌های جدید اخذ شده بر اجرای عملیات حسابرسی و پوشش مناسب مخاطرات برآوردی در اظهارنظر حسابرسی دریافت گردید.

در این خصوص حسابرسان ضمن تطبیق نمونه‌های اخذ شده با مواردی که به صورت شهودی و قضاوتی نسبت به ارزیابی آن‌ها اقدام کرده بودند، بیان داشتند که نمونه‌های اخذ شده با استفاده از تکنیک‌های داده‌کاوی، در برخی موارد پوشش مناسبتری نسبت به ادعاهای مطرح شده در صورت‌های مالی داشته‌اند و در صورتیکه از نمونه‌های جدید استفاده می‌کردند، امکان اجرای آزمون‌های اضافی جهت بررسی عمیق تر و کشف روابط بیشتر در رویدادهای مالی را فراهم می‌نمودند. در این راستا از حسابرسان در خواست شد براساس میزان پیش بینی بهبود پوشش خطا، از طریق نمره دهی در طیفی بین صفر تا صد برای نمونه‌های جدید اقدام نمایند. خلاصه نظرات ارزیابی تطبیقی نمونه‌های حسابرسی و نمرات تخصیص یافته توسط حسابرسان به نمونه‌های جدید د رخصوص بهبود پوشش، در جدول شماره ۷ بیان شده است.

جدول ۷ تطبیق نمونه انتخاب شده از طریق داده‌کاوی با نمونه‌های سنتی

شماره مجموعه داده	درصد نمونه انتخاب شده از کل داده‌ها	درصد تطابق دو نمونه	آیا تمایل دارید نمونه قبلی را با نمونه جدید جایگزین کنید؟	درصد بهبود پوشش خطا در نمونه جدید
۱	۲۵	۷۶	بله	۸۵
۲	۳۵	۷۲	بله	۸۰
۳	۳۵	۶۸	بله	۸۰
۴	۲۵	۷۸	خیر	۶۰
۵	۴۵	۵۲	بله	۷۵
۶	۴۰	۶۳	خیر	۶۰
۷	۳۵	۶۱	خیر	۶۰
۸	۳۰	۵۸	بله	۹۰
۹	۲۵	۷۹	بله	۸۰
۱۰	۳۰	۸۲	بله	۸۵
۱۱	۳۵	۸۰	بله	۸۰
۱۲	۴۵	۷۳	خیر	۶۰
۱۳	۲۰	۷۲	خیر	۵۰
۱۴	۲۵	۵۹	بله	۷۵
۱۵	۳۰	۶۷	بله	۷۵
۱۶	۳۵	۷۲	بله	۸۵
۱۷	۳۰	۷۹	بله	۸۰
۱۸	۳۰	۷۱	خیر	۵۰
۱۹	۴۰	۷۵	بله	۷۵
۲۰	۳۵	۷۷	بله	۸۰
۲۱	۲۵	۷۳	خیر	۶۵
۲۲	۲۵	۶۵	خیر	۶۵
۲۳	۳۵	۶۸	خیر	۵۵
۲۴	۳۵	۷۴	بله	۸۵
۲۵	۴۰	۶۱	بله	۸۰
۲۶	۴۵	۶۶	بله	۹۰
۲۷	۲۵	۵۶	خیر	۵۵
۲۸	۳۰	۶۳	بله	۸۵
۲۹	۳۵	۵۹	بله	۸۰
۳۰	۲۰	۶۷	بله	۸۵
۳۱	۲۵	۶۴	خیر	۶۵
۳۲	۲۰	۷۸	بله	۸۰
۳۳	۳۵	۷۱	بله	۹۰
۳۴	۴۰	۵۴	خیر	۵۵
۳۵	۴۰	۸۱	خیر	۵۰

۵- بحث و نتیجه‌گیری

آنچه در این پژوهش بررسی شده است، تعریف الگویی برای انتخاب نمونه از طریق تکنیک‌های داده‌کاوی است. براساس نتایج این پژوهش، حسابرسان می توانند با استفاده از تکنیک های داده‌کاوی نسبت به انتخاب نمونه مفید و مؤثر در فرآیند حسابرسی، اقدام نمایند.

براساس مصاحبه‌های صورت گرفته با حسابرسان و بررسی پرونده‌های حسابرسی مختلف، نویسندگان این مقاله بدین قضاوت رسیده‌اند که احتمالا تعداد قابل توجهی از حسابرسان اطلاعات دقیقی از داده‌کاوی و تکنیک‌های کاربردی آن ندارند. برداشت نویسندگان این است که عدم وجود شناخت کافی در این زمینه، باعث شده است که حسابرسان نسبت به استفاده از داده‌کاوی و تکنیک‌های آن برای انتخاب نمونه حسابرسی محتاط بوده و نگران هستند نمونه‌های انتخاب شده برای رسیدن به اهداف حسابرسی، مناسب نباشند.

از طرف دیگر ایرادات وارد بر استفاده از روش‌های سنتی نمونه‌گیری، همواره ریسک عدم انتخاب نمونه صحیح یا انتخاب نمونه‌ها براساس یک قضاوت اشتباه را افزایش می‌دهد که این موضوع باعث کاهش اثربخشی رسیدگی حسابرسان براساس نمونه‌های انتخابی خواهد بود. همچنین زمان انتخاب نمونه‌ها از طریق تکنیک‌های داده‌کاوی می‌تواند به میزان قابل توجهی کاهش یابد. زیرا تکنیک‌های داده‌کاوی می‌توانند با خوشه بندی مناسب رویدادها و تراکنش‌های مالی ثبت شده، دانش‌های نهفته در روابط داده‌ها را کشف کنند و سرعت انتخاب یک نمونه مؤثر و کارا را بهبود بخشند.

نتایج حاصل از ارزیابی اطلاعات و بررسی داده‌ها نشان می‌دهد:

۱- تکنیک‌های داده‌کاوی ضمن کشف روابط بین حساب‌های معین و متقابل در ثبت‌های مالی، نمونه‌هایی را پیشنهاد دادند که در ارزیابی بهتر این روابط مورد استفاده قرار گیرند. پیشنهاد نمونه حسابرسی از میان حساب‌های معین متقابلی که در دو سمت بدهکار و بستانکار اسناد مالی از همبستگی مناسبی برخوردار هستند و شناسایی موارد عدم تبعیت از الگوی رفتاری حساب‌های متقابل، می‌تواند در بهبود اجرای فرآیند حسابرسی و شناسایی حیطه‌های خطر احتمالی در ثبت رویدادها، مؤثر باشد.

۲- نمونه حسابرسی انتخاب شده از طریق تکنیک‌های داده‌کاوی در ۲۱ پایگاه داده، بیش از ۶۵ درصد با نمونه‌های قبلی مطابقت داشت. همچنین، در ۲۲ مورد، حسابرسان نمونه انتخاب شده با استفاده از تکنیک‌های داده‌کاوی را با نمونه‌های قبلی جایگزین کرده‌اند. از موارد فوق‌الذکر، ۱۵ مورد مربوط به نمونه‌هایی بود که بیش از ۶۵درصد با نمونه‌های قبلی مطابقت داشتند. بر این اساس نویسندگان استدلال می‌کنند که تکنیک‌های داده‌کاوی، نمونه‌های مناسبی را برای استدلال منطقی در فرآیند اجرای حسابرسی، به حسابرسان ارایه داده‌اند.

۳- با استفاده از تکنیک‌های داده‌کاوی، میانگین زمان انتخاب نمونه برای هر دسته از اطلاعات کمتر از ۱۵ دقیقه است، در حالی که با روش‌های سنتی، به طور متوسط بیش از سه روز زمان

برای انتخاب نمونه در بخش‌های مختلف صرف می‌شود. بنابراین، به نظر می‌رسد که کاربرد داده‌کاوی انتخاب نمونه‌ها را تسریع کرده و موجب بهبود کارایی فعالیت‌های حسابرسی شود.

نکته قابل توجه در این پژوهش آن است که تاکنون پژوهشی در خصوص انتخاب نمونه‌های حسابرسی با استفاده از تکنیک‌های داده‌کاوی صورت نپذیرفته است. این رویکرد نوآورانه از یک طرف می‌تواند نقطه آغازی برای استفاده از تکنیک‌های داده‌کاوی در حسابرسی باشد و از طرف دیگر می‌تواند در سایر بخش‌های حسابرسی از جمله ارزیابی ریسک حسابرسی مورد استفاده قرار گیرد. بدیهی است توسعه فعالیت‌های پژوهشی در این حوزه می‌تواند به تعمیق بیشتر نتایج آن و انتخاب تکنیک‌های داده‌کاوی مؤثرتر، منجر شود.

یکی از مهمترین محدودیت‌های اجرای این پژوهش، دسترسی به پایگاه‌های داده شرکت‌های مختلف بود. از آنجائی‌که اطلاعات مالی شرکت‌ها، جز اطلاعات محرمانه طبقه‌بندی می‌شود، دسترسی به این داده‌ها برای انجام پژوهش‌های مشابه با محدودیت‌های جدی روبرو است. پایگاه‌های داده انتخاب شده در این پژوهش صرفا مربوط به پروژه‌های حسابرسی یک مؤسسه حسابرسی است و دسترسی به اطلاعات و پایگاه‌های داده سایر شرکت‌های تحت حسابرسی مؤسسات دیگر، می‌تواند به تقویت نتیجه‌گیری این پژوهش کمک نماید. محدودیت دیگر این پژوهش آن است که عدم وجود دانش کافی و مناسب در میان حسابرسان، استفاده از تکنیک‌های داده‌کاوی برای بخش‌های مختلف حسابرسی را با چالش مواجه می‌کند.

به نظر نویسندگان، اجرای پژوهش‌های آتی در حسابرسی می‌تواند در دو بخش ادامه یابد:

بخش اول استفاده از تکنیک‌های مختلف و تکمیلی در ارزیابی یک بخش از حسابرسی است؛ به عنوان مثال از سایر تکنیک‌های خوشه بندی نیز برای ارزیابی نمونه‌های حسابرسی استفاده شده و نتایج آن با تحقیق حاضر مورد مقایسه قرار گیرد.

بخش دوم استفاده از تکنیک‌های داده‌کاوی در بخش‌های دیگر حسابرسی است؛ به عنوان مثال استفاده از تکنیک‌های درخت تصمیم در ارزیابی خطر عدم کشف حسابرسی می‌تواند به عنوان یک پژوهش آتی در این زمینه مطرح شود.

همچنین نویسندگان معتقدند تکنیک‌های داده‌کاوی به لحاظ ارزیابی، دقیق‌تر و بدون قضاوت و جانبداری هستند و می‌توانند در انتخاب یک نمونه مؤثر، مناسب و دقیق در اجرای آزمون‌ها و رسیدگی‌های حسابرسی مورد استفاده قرار گیرند. تکنیک‌های داده‌کاوی با تسریع عملیات انتخاب نمونه مناسب، کارایی عملیات حسابرسی را بهبود می‌بخشند.

منابع

علوی، سیدکمال و نعمتی، علی و دارابی، رویا (۱۴۰۴). الگوی بهینه و کاربردی فناوری اطلاعات در حسابرسی با توجه به آزمون‌های محتوا، کنترل و ریسک‌های حسابرسی، پژوهش‌های حرفه ای حسابرسی، ۵(۲۰)، ۳۰-۵۷.

معطوفی، علیرضا و شیخ عبدالکریم ، فریال وگرکز،منصور و خوزین، علی. (۱۴۰۳). شناسایی

عوامل مؤثر بر انگیزه رفتارهای زورگویانه حسابرسان نسبت به یکدیگر، پژوهش‌های حرفه ای حسابرسی، ۵(۱۷)، ۸-۳۵.

Alavi, S. K., Nemati, A., & Darabi, R., (2025). An Optimal and Practical Model of Information Technology in Auditing Considering Substantive and Control Testing, and Audit Risk, Journal of Professional Auditing Research, Vol. 5, No. 20, 30-57. (in persian). https://doi.org/10.22034/jpar.2024.2031862.1327

Amani, F. A., & Fadlalla, A., (2017). Data mining applications in accounting: A review of the literature and organizing framework, International Journal of Accounting Information Systems., vol. 24, no.2, 32–58.

https://www.researchgate.net/publication/312961430_Data_mining_applications_in_accounting_A_review_of_the_literature_and_organizing_framework

Arens, A. A., Elder, R. J., & Beasley, M. S., (2014). Auditing and Assurance Services: An Integrated Approach., 2nd ed, Pearson.

Awad, S. S., & Wathik, I. M. (2022). Using Data Mining Tools to Predict Going Concern on Auditor Opinion. Academy of Accounting and Financial Studies Journal, 26 (S23), 1-13.

https://www.researchgate.net/publication/358802230_Using-Data-mining-tools-to-Prediction-of-going-Concern-on-Auditor-Opinion-Empirical-study-In-Iraqi-Commercial-1528-2635-26-3-1010

Arisudhana, A., & Rohmah, K. L. (2023). Data Mining in Auditing: Challenges and Opportunities. International Conference on Information Science and Technology Innovation (ICoSTEC) 2(1): 178-180

https://www.researchgate.net/publication/372893263_Data_Mining_in_Auditing_Challenges_and_Opportunities

Brown, C., & Vasarhelyi, M. A., (2019). Continuous auditing: A new view. J. Emerging Technol. Account., vol. 16, no. 2, 1–10. https://doi.org/10.1108/978-1-78743-413-420181002

Cascarino, R. E., (2012). Auditor's Guide to IT Auditing. 1st ed, John Wiley & Sons.

Deloitte (2024). How WestRock Harnessed GenAI to Enhance Internal Audit. https://deloitte.wsj.com/riskandcompliance/how-westrock-harnessed-genai-to-enhance-internal-audit-f0926363

Elder, R. J., & Allen, R. D., (2000). An Empirical Investigation of the Relation Between Risk Assessments and Sample Size Decisions, SSRN Electronic Journal., https://papers.ssrn.com/sol3/papers.cfm?abstract_id=211848

Elder, R. J., Akresh, A. D., Glover, S. M., Higgs, L. J., & Liljegren, J., (2013) Audit Sampling Research: A Synthesis and Implications for Future Research, A Journal of Practice & Theory, vol. 32, no. 1, 99–129.

https://doi.org/10.2308/ajpt-50394

Gupta, R., (2019). Data Mining for Fraud Detection: An Overview of Techniques and Applications, Turkish Journal of Computer and Mathematics Education, vol. 10, no. 1, 561–567.

https://doi.org/10.17762/turcomat.v10i1.13549

Han, J., Kamber, M., & Pei, J., (2012). Data Mining: Concepts and Techniques, Morgan Kaufmann, 3rd edition.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters. Vol. 31, No, 8, 651-666.

https://doi.org/10.1016/j.patrec.2009.09.011

Knechel, W. R., Salterio, S. E. & Ballou, B., (2007). Auditing: Assurance and Risk. Thomson South-Western. https://doi.org/10.4324/9781315531731

Pycka, M., & Zastempowski, M. (2025). Machine Learning and Artificial Intelligence Techniques Adopted for IT Audit, Journal of Management, vol 29, No. 1, 65-87. https://doi.org/10.58691/man/200768

Santoso, F., Wulandari, I., & Partiwi, D., (2023). Evaluation of Sampling Techniques in Audit: A Qualitative Approach. Golden Ratio of Auditing Research, 3(1), 11–20. https://doi.org/10.52970/grar.v3i1.373

Sheikhabdolkarim, F., Matoufi, A., Garkaz, M., & Khoazain, A, (2024). Identifying Factors Influencing Motivation of The Auditors' Bullying Behaviors --Towards Each Other, Journal of Professional Auditing Research, Vol. 5, No. 17, 8-35. (in persian)

https://doi.org/10.22034/JPAR.2024.2016879.1253

Sheu, G.-Y., & Liu, N.-R. (2024). Sampling Audit Evidence Using a Naive Bayes Classifier. https://arxiv.org/abs/2403.14069.

Witten, I. H., Frank, E. & Hall, M. A., (2011). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

پی‌نویس:

1. Knechel, Salterio, & Ballou
2. Pycka & Zastempowski
3. Deloitte
4. Han, Kamber & Pei
5. Arens, Elder, & Beasley
6. Elder and Allen
7. Elder & et al
8. Santoso, Wulandari & Partiwi,
9. Arens, Elder, & Beasley
10. Elder and Allen
11. Knechel & et al
12. Gupta
13. Amani & Fadlalla
14. Koh & Tan
15. Brown & Vasarhelyi
16. Cascarino
17. Sheu & Liu
18. Jane
19. Bremen
20. Avad & vatik
21. pastor