



## Application of Data Mining in Tax Processes Improvement: A Literature Review and Classification

### Reza Majidi

Ph.D. Candidate, Department of Accounting, Faculty of Management and Accounting, Center Tehran Branch, Islamic Azad University, Tehran, Iran. E-mail: reza1180@gmail.com

### Negar Khosravipour\*

\*Corresponding Author, Assistant Prof., Department of Accounting, Faculty of Management and Accounting, Center Tehran Branch, Islamic Azad University, Tehran, Iran. E-mail: n\_khosravipour@yahoo.com

### Elham Akhondzadeh Noughabi

Assistant Prof., Department of IT Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran. E-mail: elham.akhondzadeh@modares.ac.ir

### Abstract

**Objective:** Data mining is an effective tool to improve and enhance the efficiency and effectiveness of tax processes by extracting beneficial knowledge and insight from tax data. The purpose of this paper is to study the status of research pieces in the field, classify them, identify the research gaps and provide a roadmap for researchers and practitioners through a systematic literature review.

**Methods:** Due to the importance of the subject, this study focuses on studies conducted in the field of data mining and taxation from 2000 to 2021. It investigates their processes and practical domains. The reviewed studies were categorized based on the proposed framework and various analyses were presented in terms of processes, practical domains, and data mining techniques. Furthermore, the distribution of papers according to the year of their publication and also regarding the journal in which they were published were presented. Tax processes were divided into four groups i.e. submission, examination, collection, and taxpayer services. The defined practical domains included tax payment, tax refund, shell corporation identification, identification of non-filer taxpayers, risk-based tax audit selection, tax debt management, and tax comments analysis. The classification framework for data mining techniques in this study was defined as clustering, association analysis, classification and prediction, regression, time series, anomaly detection, and visualization.

**Results:** According to the obtained findings, most of the reviewed studies were assigned to the inspection process, 94 percent of which worked on the practical domain of “risk-based tax audit selection”. The most popular and widely used technique was "classification and prediction", while the three algorithms including neural network, decision tree, and support vector machine were widely used, compared to other algorithms.

**Conclusion:** Currently, tax administrations have huge databases and traditional methods and tools cannot analyze them due to the limited resources of organizations as well as the large amounts of available data. Data mining can have an effective performance on various tax processes and can be effective in making decisions and adopting appropriate approaches. There is good potential for the application of data mining techniques in all of the proposed practical domains. In the submission and collection processes, more research needs to be done. Some approaches including reinforcement learning, deep learning, graph analysis, and big data analytics are recommended for future research. Proposing practical frameworks for using data mining techniques in tax systems and tax administrations is also recommended. To the best of the author's knowledge, no study has been conducted to investigate the issue, while there is a definite need in this regard. Besides, one of the important issues, which needs to be addressed as the main gap in this field, is integrating the internal and external sources of data, which can improve effectiveness.

**Keywords:** Data mining, Classification and prediction, Tax process, Taxation.

**Citation:** Majidi, Reza; Khosravipour, Negar & Akhondzadeh Noughabi, Elham (2022). Application of Data Mining in Tax Processes Improvement: A Literature Review and Classification. *Accounting and Auditing Review*, 29(3), 519-545. (in Persian)

---

Accounting and Auditing Review, 2022, Vol. 29, No.3, pp. 519- 545  
Published by University of Tehran, Faculty of Management  
<https://doi.org/10.22059/ACCTGREV.2022.331897.1008602>  
Article Type: Research Paper  
© Authors

Received: November 06, 2022  
Received in revised form: February 17, 2022  
Accepted: February 21, 2022  
Published online: October 22, 2022





## کاربرد داده‌کاوی در بهبود فرایندهای مالیاتی: مرور ادبیات سیستماتیک و دسته‌بندی

رضا مجیدی

دانشجوی دکتری، گروه حسابداری، دانشکده مدیریت و حسابداری، واحد تهران مرکز، دانشگاه آزاد اسلامی تهران، ایران. رایانامه: reza1180@gmail.com

نگار خسروی‌پور\*

\* نویسنده مسئول، استادیار، گروه حسابداری، دانشکده مدیریت و حسابداری، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران. رایانامه: n\_khosravipour@yahoo.com

الهام آخوندزاده نوقابی

استادیار، گروه مهندسی فناوری اطلاعات، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران. رایانامه: elham.akhondzadeh@modares.ac.ir

### چکیده

**هدف:** با توجه به اهمیت استخراج دانش مفید از داده‌های مالیاتی و نقش مؤثر داده‌کاوی در این زمینه، هدف این پژوهش، مرور ادبیات جامع و نظام‌مند و ارائه گزارشی از وضعیت تحقیقات حوزه داده‌کاوی و مالیات، دسته‌بندی پژوهش‌های انجام‌شده و معرفی شکاف‌های تحقیقاتی و ارائه نقشه راهی برای محققان و علاقه‌مندان در این زمینه است.

**روش:** جامعه آماری پژوهش، تحقیقات انجام‌شده در زمینه داده‌کاوی و مالیات، طی سال‌های ۲۰۰۰ تا ۲۰۲۱ بوده است. با مرور ادبیات جامع و نظام‌مند تحقیقات، از دیدگاه فرایندی، ۴ فرایند و از دیدگاه حوزه‌های کاربردی مختلف، ۷ زمینه مطالعه و بررسی شد. دسته‌بندی تحقیقات بر اساس چارچوب پیشنهادی انجام گرفت و تحلیل‌های مختلفی از منظر فرایندها، حوزه‌های کاربردی و روش‌های داده‌کاوی ارائه شد.

**یافته‌ها:** نتایج این تحقیق نشان می‌دهد که فرایند بازرسی (آزمون) و حوزه کاربردی «انتخاب مبتنی بر ریسک برای حسابرسی مالیاتی» بیشترین حجم از تحقیقات را به خود اختصاص داده‌اند. محبوب‌ترین و پرکاربردترین تکنیک استفاده شده، «رده‌بندی و پیش‌بینی» بوده و الگوریتم‌های شبکه عصبی، درخت تصمیم و ماشین بردار پشتیبان نیز به ترتیب بیشترین کاربرد را داشته‌اند.

**نتیجه‌گیری:** در زمینه‌های کاربردی هفت‌گانه، پتانسیل خوبی برای پیاده‌سازی تکنیک‌های داده‌کاوی وجود دارد. رویکردهای مبتنی بر یادگیری انتقالی، یادگیری عمیق، تحلیل گراف و تحلیل کلان‌داده برای تحقیقات آتی پیشنهاد می‌شود. ارائه چارچوب‌های کاربردی بومی‌سازی شده برای سیستم‌ها و اداره‌های امور مالیاتی کشورهای مختلف و یکپارچه‌سازی منابع داده داخلی و خارجی اداره‌های امور مالیاتی و تحلیل آن، از خلأهای اصلی این حوزه است که می‌تواند اثربخشی ویژه‌ای ایجاد کند.

**کلیدواژه‌ها:** داده‌کاوی، رده‌بندی و پیش‌بینی، فرایند مالیاتی، مالیات.

**استناد:** مجیدی، رضا؛ خسروی‌پور، نگار و آخوندزاده نوقابی، الهام (۱۴۰۱). کاربرد داده‌کاوی در بهبود فرایندهای مالیاتی: مرور ادبیات سیستماتیک و دسته‌بندی. *بررسی‌های حسابداری و حسابرسی*، ۲۹(۳)، ۵۱۹-۵۴۵.

## مقدمه

یکی از مؤلفه‌های اصلی سیستم اقتصادی هر کشوری، مالیات<sup>۱</sup> است که مهم‌ترین ابزار کسب درآمد دولت‌ها برای انجام فعالیت‌های اقتصادی مختلف محسوب می‌شود. برای همگان روشن است که موضوع مالیات و نقش نظام مالیاتی در وصول مالیات، از جایگاه ویژه‌ای برخوردار است. دولت‌ها با دریافت انواع مالیات از افراد حقیقی و حقوقی، درآمد لازم را برای انجام وظایف خود تأمین می‌کنند. مالیات روشی برای تأمین مخارج سرمایه‌ای و سرمایه در گردش دولت بوده و از سوی دیگر، استقرار نظام مطلوب و کارایی مالیاتی، توزیع صحیح ثروت و درآمد در جامعه را به دنبال دارد (مام موسی اگریفاش، ۱۳۹۴ و نمازیان، پورحیدری و زینلی، ۱۴۰۰).

یکی از تکنولوژی‌هایی که می‌تواند در راستای تحقق نظام مطلوب و کارایی مالیاتی اثربخش باشد، داده‌کاوی<sup>۲</sup> است. تکنیک‌های تحلیل داده و داده‌کاوی می‌تواند در کشف فرار مالیاتی<sup>۳</sup> و حوزه‌های عملکردی مختلف دیگر مؤثر واقع شود و از این‌رو مورد توجه محققان و علاقه‌مندان دنیای واقعی قرار گرفته است. تمایل سازمان‌های مالیاتی برای به‌کارگیری تکنیک‌های داده‌کاوی به‌طور چشمگیری افزایش یافته و سازمان‌های بیشتری به کاربرد داده‌کاوی در حل مسائل مالیاتی روی می‌آورند (بایسالباویا، ون‌دراندن، تنان و فلارس<sup>۴</sup>، ۲۰۱۸).

با توجه به اهمیت و نقش ویژه کاربردهای داده‌کاوی در حوزه مالیات، در این پژوهش کوشش شده تا یک مرور ادبیات سیستماتیک و نظام‌مند در خصوص تحقیقات موجود در این زمینه انجام شود؛ همچنین چارچوب و طراحی برای دسته‌بندی تحقیقات ارائه می‌گردد. پژوهش‌های مختلف در این حوزه مطالعه و دسته‌بندی شده و تلاش می‌گردد با یک رویکرد جامع‌نگر و دیدگاه نقادانه، حوزه‌های پرکاربرد و همین‌طور شکاف‌های تحقیقاتی مشخص شوند. کاربردهای داده‌کاوی در حوزه مالیات در سطح اول از دیدگاه فرایندی و در سطح دوم از منظر حوزه‌های کاربردی مختلف در هر یک از فرایندها بحث می‌شود. در نهایت، تحقیقات مورد بررسی بر اساس چارچوب ارائه شده دسته‌بندی شده و شکاف‌های تحقیقاتی و پیشنهادهایی برای پژوهش‌های آتی ارائه می‌شود. بر اساس بررسی‌های انجام شده، پژوهش حاضر اولین موردی است که با رویکرد جامع‌نگر و نظام‌مند به مرور ادبیات در حوزه کاربردهای داده‌کاوی در مالیات با دیدگاه فرایندی می‌پردازد و چارچوبی برای دسته‌بندی تحقیقات ارائه می‌کند. با توجه به دسته‌بندی فرایندی و کاربردی، این مقاله می‌تواند توسط سازمان‌های مالیاتی به‌صورت کاربردی نیز مورد استفاده قرار گیرد. تنها پژوهش مرتبط انجام شده در این زمینه، اردونز و هالو<sup>۵</sup> (۲۰۱۹) می‌باشد که به مرور ادبیات کاربردهای داده‌کاوی در دستگاه‌های مالیات‌ستان پرداخته است. تفاوت تحقیق حاضر با اردونز و هالو (۲۰۱۹) در این است که ادبیات موجود و تحقیقات انجام شده را به‌صورت بسیار جامع‌تر بررسی نموده و با دیدگاه فرایندی، یک چارچوب هدفمند برای دسته‌بندی تحقیقات در دو سطح فرایندی و کاربردی ارائه می‌دهد. پژوهش اردونز و هالو (۲۰۱۹) رویکرد جامع‌نگری نداشته؛ به لحاظ مقالات و تحقیقات بررسی شده محدودتر است و تعداد کمی را مورد بررسی قرار داده است. همچنین پژوهش حاضر علاوه بر حوزه‌های بررسی شده

1. Tax
2. Data Mining
3. Tax Evasion
4. Baisalbayeva, van der Enden, Tenan & Flores
5. Ordóñez & Hallo

در تحقیق مذکور، حوزه‌های کاربردی شناسایی مؤدیان فاقد اظهارنامه، شناسایی مؤدیان صوری (کد فروش)، استرداد مالیات و تحلیل نظرات مالیاتی را نیز پوشش می‌دهد. تکنیک‌های داده‌کاوی مورد بررسی در پژوهش مذکور نیز محدودتر و شامل تنها رده‌بندی<sup>۱</sup>، خوشه‌بندی<sup>۲</sup>، خلاصه‌سازی<sup>۳</sup> و پیش‌بینی<sup>۴</sup> است؛ در حالی که تحقیق حاضر تحلیل انجمنی<sup>۵</sup>، رگرسیون<sup>۶</sup>، سری‌های زمانی<sup>۷</sup>، تشخیص ناهنجاری<sup>۸</sup> و مصورسازی<sup>۹</sup> را نیز در نظر گرفته است. عدم نقد ادبیات، شناسایی خلأهای تحقیقاتی و پیشنهادات برای تحقیقات آتی نیز از نقاط ضعف پژوهش مذکور است.

ساختار مقاله به این صورت است: در بخش آتی روش‌شناسی پژوهش و چارچوب پیشنهادی برای دسته‌بندی مقالات ارائه می‌گردد. سپس پیشینه پژوهش ارائه می‌شود. در ادامه مقالات بر اساس چارچوب پیشنهادی، دسته‌بندی شده و نتایج آن تحلیل می‌گردد؛ همچنین به توزیع مقالات بر اساس روش داده‌کاوی، حوزه کاربردی، سال انتشار و نشریه منتشر شده می‌پردازیم. در نهایت در بخش پایانی، نتیجه‌گیری و پیشنهادات بیان می‌شود.

### روش‌شناسی پژوهش

در این تحقیق، مطالعه جامع و نظام‌مندی از پژوهش‌هایی که در زمینه کاربردهای تکنیک‌های داده‌کاوی در حوزه مالیات انجام شده، صورت می‌گیرد. به عبارت دقیق‌تر، این پژوهش به دنبال پاسخ به پرسش‌های زیر است:

- آیا از داده‌کاوی در حوزه مالیات استفاده شده است؟
- کاربردهای داده‌کاوی چه فرایندهای مالیاتی و حوزه‌های کاربردی را شامل می‌شود؟
- از داده‌کاوی به چه صورت می‌توان در فرایندهای مختلف مالیاتی استفاده کرد؟
- کدام تکنیک‌های داده‌کاوی بیشتر در حوزه مالیات مورد استفاده قرار گرفته‌اند؟
- داده‌کاوی در کدام فرایندهای مالیاتی و حوزه‌های کاربردی، بیشتر استفاده شده و اثربخش بوده است؟
- شکاف‌ها، فرصت‌ها و محدودیت‌های تحقیقاتی این حوزه چیست؟

در راستای پاسخ‌گویی به پرسش‌های مورد اشاره، به جست‌وجوی مقالات از طریق پایگاه‌های معتبر پرداخته شد. در ابتدا جست‌وجوی انجام شده با دو کلیدواژه «مالیات» و «داده‌کاوی» صورت گرفت. سپس کلیدواژه‌های دیگر مرتبط با داده‌کاوی مانند خوشه‌بندی، تحلیل انجمنی، رگرسیون، رده‌بندی، پیش‌بینی، سری‌های زمانی، کاوش دنباله‌ای<sup>۱۰</sup>، تشخیص ناهنجاری و مصورسازی به همراه مالیات، فرایند مالیاتی<sup>۱۱</sup>، مالیات بر ارزش افزوده<sup>۱۲</sup>، تشخیص فریب<sup>۱۳</sup>، فرار

- 
1. Classification
  2. Clustering
  3. Summarization
  4. Prediction
  5. Association Analysis
  6. Regression
  7. Time Series
  8. Anomaly Detection
  9. Visualization
  10. Sequence Mining
  11. Tax Process
  12. Value Added Tax
  13. Fraud Detection

مالیاتی، اجتناب مالیاتی<sup>۱</sup> و همین طور اداره‌های مالیاتی<sup>۲</sup> نیز جست‌وجو شده‌اند. پایگاه داده‌های برخطی که جست‌وجو در آن‌ها انجام شده است، عبارت‌اند از: ScienceDirect، IEEE xplore digital library (IEEE Xplore)، ACM Digital Library و Scopus، ProQuest، Emerald، SpringerLink مربوط به نشریه‌ها و کنفرانس‌های بین‌المللی معتبر مبنای مطالعه قرار گیرد. شایان ذکر است فرایند جست‌وجوی مقالات و انتخاب مقاله، توسط دو محقق به صورت مستقل و موازی انجام شده و مورد بررسی قرار گرفته است. آن دسته از مقالاتی که مرتبط با کاربرد تکنیک‌های داده‌کاوی در فرایندهای مالیاتی نبودند، از مطالعه حذف شده‌اند. در ابتدا حدود ۱۰۰ مقاله بررسی شده و در نهایت ۵۵ مورد، انتخاب و تجزیه و تحلیل شدند. تکنیک‌ها و الگوریتم‌های مورد استفاده و فرایند مالیاتی مرتبط در هر مقاله، توسط دو محقق مشخص و مورد بازبینی قرار گرفت و بر این اساس نهایی شد. در ادامه این مقالات از نظر تکنیک و الگوریتم داده‌کاوی مورد استفاده در فرایندهای مختلف سازمان‌های مالیاتی و حوزه‌های کاربردی آن، دسته‌بندی شدند و تحلیل‌های مختلفی روی آن صورت گرفت که در بخش‌های آتی ارائه می‌گردد. مقالات بر اساس معیارهای زیر انتخاب شده‌اند:

- تنها آن دسته از مقالاتی که مشخصاً مرتبط با کاربردهای داده‌کاوی در حوزه فرایندهای مالیاتی بوده و به‌طور کامل پیاده‌سازی الگوریتم‌ها و روش‌های داده‌کاوی را در حوزه‌ای مربوط به فرایندهای مالیاتی شرح داده بودند، انتخاب شده‌اند. در انتخاب مقالات به وجود مشارکت کاربردی در بهبود فرایندهای مالیاتی و یا نوآوری تئوری در روش توجه شده و این گونه مقاله‌ها در تحلیل‌ها در نظر گرفته شده‌اند.
- تنها مقاله‌هایی که به زبان انگلیسی در دسترس بودند، انتخاب شدند.
- فقط مقالات ژورنال‌ها و کنفرانس‌ها برگزیده شدند. البته کنفرانس‌ها محدود به کنفرانس‌های بین‌المللی معتبر به ویژه IEEE بوده است.

در نهایت با توجه به مطالعات انجام شده، چارچوبی برای دسته‌بندی مقالات مبتنی بر فرایندهای مالیاتی و حوزه‌های کاربردی مختلف در هر یک از فرایندها، ارائه شده و توسط سه خبره از سازمان امور مالیاتی ایران نیز مورد تأیید قرار گرفته است. تمامی مقالات انتخابی نهایی بر اساس چارچوب ارائه شده، دسته‌بندی و تحلیل‌های متنوعی از منظرهای مختلف انجام شده که نتایج آن در بخش‌های آتی ارائه می‌گردد. در ادامه چارچوب پیشنهادی ارائه می‌گردد.

### چارچوب پیشنهادی دسته‌بندی تحقیقات

مالیات جزئی ضروری از جامعه مدرن امروزی است؛ موضوع مالیات‌ستانی تبدیل به یکی از موضوعات مهم در دنیا و ایران شده و آحاد جامعه سعی در فهم چگونگی کارکرد آن دارند. مسئولیت اولیه سیستم مالیاتی در هر کشوری، وصول قابل قبول درآمدهای مالیاتی با حداقل هزینه ممکن برای جامعه می‌باشد. از طرفی انجام این مسئولیت‌ها بایستی به‌گونه‌ای باشد که بیشترین اعتماد را از سوی جامعه ایجاد نماید تا عملاً تمامی افراد به پرداخت مالیات تمکین نمایند. سیستم مالیاتی هر کشوری در واقع به این سؤال پاسخ می‌دهد که سازمان‌های مالیاتی برای انجام مسئولیت‌های خود،

نیاز به چه فرایندهایی دارند (پیستون، رالولد، هتینگ، نگیوایرا و وست<sup>۱</sup>، ۲۰۱۹). با توجه به اهمیت فرایندهای مالیاتی، در این پژوهش دسته‌بندی فرایندهای مالیاتی بر اساس مفهوم تمکین مالیاتی و منطبق با اهداف اصلی فرایندهای مالیاتی انجام شده و به‌عنوان چارچوب حوزه مالیات برای دسته‌بندی تحقیقات در نظر گرفته شده است.

کارایی و اثربخشی در وصول مالیات هدف اصلی کلیه فرایندهای مالیاتی می‌باشد. اثربخشی و کارایی به میزان نتایج بدست آمده و ارتباط آن با منابع مصرف شده از جمله هزینه‌ها و زمان اشاره داشته که از طریق وصول قابل قبول درآمدهای مالیاتی با حداقل هزینه ممکن محقق می‌گردد. تعیین میزان کارایی و اثربخشی در سطح فرایندهای مالیاتی با اندازه‌گیری سطح تمکین در کلیه فرایندهای مالیاتی امکان‌پذیر می‌باشد. به عبارت دیگر دسته‌بندی انجام شده بر اساس تمکین مالیاتی کاملاً با اهداف فرایندهای مالیاتی مطابقت دارد. از طرفی داده‌کاوی از طریق تحلیل داده‌های مالیاتی، ابزاری مفید جهت ارتقاء کارایی و اثربخشی فرایندهای مالیاتی محسوب شده و می‌تواند به مدیران مالیاتی در تصمیم‌گیری‌های کوتاه‌مدت و بلندمدت آنان کمک کند.

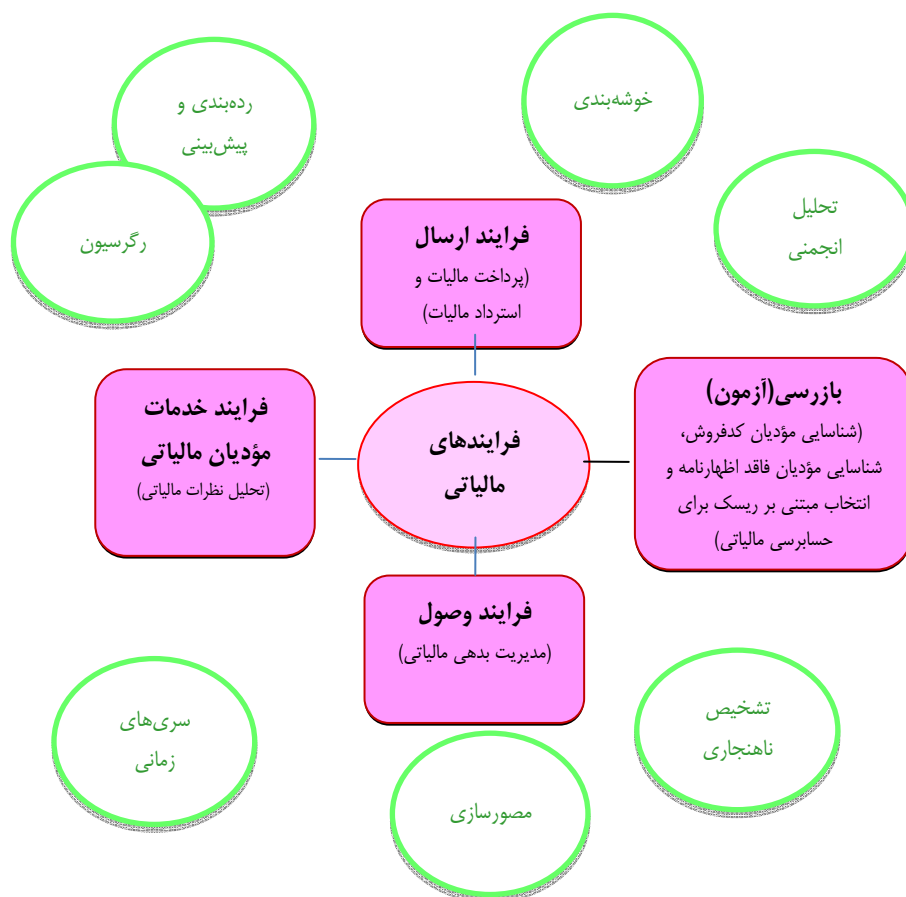
تعیین سطح تمکین مؤدیان برای سازمان امور مالیاتی همانند اندازه‌گیری سود خالص در بخش خصوصی است. زیرا هر دو شاخص‌های خلاصه شده از کارایی و اثربخشی عملیات یک سازمان هستند. همان‌طور که محاسبه سود خالص، بیانگر عملکرد مدیریت واحد تجاری است، اندازه‌گیری تمکین نیز به نوعی نشان‌دهنده عملکرد مناسب یا نامناسب دستگاه اجرایی مالیات به‌عنوان متولی اصلی بهبود فرایندهای مالیاتی است (باباجانی و باقری، ۱۳۹۶). از سوی دیگر، تعیین حوزه‌هایی که در آنها تمکین مالیاتی کم بوده و یا عدم تمکین وجود دارد، در کنار رتبه‌بندی این حوزه‌ها، به‌عنوان بخشی از فرایند تخصیص منابع کمیاب به‌منظور دستیابی به بیشترین کارایی و اثربخشی محسوب می‌شود (سازمان همکاری و توسعه اقتصادی<sup>۲</sup>، ۲۰۱۰). تمکین یا عدم تمکین مالیاتی می‌تواند در تمامی فرایندهای دستگاه اجرایی تعریف گردد. در این پژوهش از تعریف ارائه شده توسط دهرمن و پینشاو (۲۰۰۹) استفاده شده که فرایندهای مالیاتی از منظر بهبود سطح تمکین به چهار فرایند ارسال<sup>۳</sup>، بازرسی<sup>۴</sup> (آزمون)، وصول<sup>۵</sup> و خدمات مؤدیان<sup>۶</sup> تقسیم‌بندی شده‌اند. شمای کلی چارچوب پیشنهادی برای دسته‌بندی تحقیقات در شکل ۱ نشان داده شده که در ادامه شرح داده می‌شود. در هر فرایند، حوزه‌های کاربردی مرتبط داخل پرانتز ذکر شده است.

### چارچوب فرایندهای مالیاتی

چارچوب فرایندهای مالیاتی شامل چهار فرایند و حوزه‌های کاربردی هر یک به شرح ذیل است. **فرایند ارسال:** این فرایند شامل تمامی فعالیت‌های معمول پرداخت مالیات و استرداد مالیات و به عبارتی شامل پرداخت مالیات به همراه تسلیم اظهارنامه مالیاتی، پرداخت‌های تکلیفی مالیات و سایر پرداخت‌ها می‌شود. تعیین مالیات قابل پرداخت در مرحله بازرسی از جمله حسابرسی و یا فرایند وصول مالیات از مؤدیان فاقد اظهارنامه در این فرایند دسته‌بندی

1. Pistone, Roeleveld, Hattingh, Nogueira & West  
 2. OECD  
 3. Submission  
 4. Examination  
 5. Collection  
 6. Taxpayer Service

نمی‌گردد. با توجه به تعریف انجام شده دو حوزه کاربردی فرایند ارسال شامل «پرداخت مالیات» و «استرداد مالیات» تعریف می‌شود. پرداخت مالیات صرفاً تا مرحله تسلیم اظهارنامه مالیاتی را شامل شده و می‌تواند شامل منابع متعدد مالیاتی مانند مالیات بر درآمد اشخاص حقوقی، مالیات بر درآمد املاک و سایر موارد باشد. پرداخت مذکور به صورت تجمعی یا به صورت منظم و ماهانه صورت گرفته و نشان‌دهنده سطح تمکین بالا از سوی مؤدیان مالیاتی است. پرداخت مالیات می‌تواند شامل پیش‌پرداخت (پرداخت قبل از تعلق مالیات)، علی‌الحساب مالیات (پرداخت مالیات توسط دیگران) و مالیات ابرازی (پرداخت به همراه اظهارنامه مالیاتی) باشد. مطابق قوانین متعدد مالیاتی در کشورهای مختلف، ممکن است مؤدیان مالیاتی مشمول استرداد مالیات گردند. استرداد مالیات به معنای عودت اضافه پرداختی مالیات به مؤدیان مالیاتی می‌باشد که از جمله موارد بسیار پرکاربرد در این حوزه می‌توان به استرداد مالیات در منبع مالیات بر ارزش افزوده اشاره نمود.



شکل ۱. چارچوب دسته‌بندی تحقیقات

**فرایند بازرسی (آزمون):** تعریف ارائه شده از فرایند بازرسی شامل بررسی موردی، حسابرسی مالیاتی جامع، محاسبه مانده مالیات قابل پرداخت مؤدیان فاقد اظهارنامه و همچنین هر برنامه اطلاع رسانی و ابتکار بازدارندگی در این خصوص



را که موجب ارتقای سطح تمکین گردد را شامل می‌شود. مؤدیان مالیاتی در فرایند بازرسی به دو دسته مؤدیان دارای اظهارنامه و فاقد اظهارنامه تقسیم‌بندی می‌گردند. یکی از راه‌کارهای بررسی سطح تمکین مؤدیان دارای اظهارنامه، انجام حسابرسی مالیاتی جامع مؤدیان مذکور است. امروزه با توجه به مشکلاتی از قبیل حجم انبوه و رو به افزایش اظهارنامه‌های مالیاتی، محدودیت زمانی و سایر عوامل، حسابرسی جامع برای تمامی مؤدیان امکان‌پذیر نبوده و لذا سازمان‌های مالیاتی اقدام به بررسی موردی اظهارنامه مالیاتی از جمله انجام حسابرسی مالیاتی بر مبنای ریسک می‌نمایند.

شناسایی مؤدیانی که موظف به تسلیم اظهارنامه مالیاتی و پرداخت مالیات بوده لیکن از انجام وظایف قانونی خودداری نموده‌اند، یکی از مشکلات عدیده دستگاه‌های اجرایی مالیات می‌باشد. تجزیه و تحلیل و استفاده از کلیه اطلاعات در دسترس برای تشخیص اینگونه مؤدیان، از جمله وظایف دستگاه مالیات‌ستان است. با توجه به اهمیت موضوع، در خصوص مؤدیان فاقد اظهارنامه که با اعمال مجرمانه و تاسیس شرکت‌های صوری (کد فروش) اقدام به فرار مالیاتی نموده‌اند، حوزه کاربردی آن به‌صورت جداگانه مورد بررسی قرار می‌گیرد. تأسیس شرکت‌های صوری و غیرواقعی به‌منظور اختفای فعالیت‌های اقتصادی، ثبت شرکت‌های صوری به‌منظور انجام اعمال مجرمانه از قبیل فرار مالیاتی، پولشویی، قاچاق کالا و یا اخذ تسهیلات بانکی به‌صورت غیرمجاز بخشی از آسیب‌های وارد به تولید و اقتصاد یک کشور می‌باشد.

**فرایند وصول:** فرایند وصول مرتبط با دریافت مانده مالیات پرداخت نشده مؤدیان متخلف از طریق مکاتبه نمودن، تلفن زدن و عملیات اجرایی می‌باشد. وصول مالیات یکی از اهداف اصلی سازمان‌های مالیاتی در دنیا بوده و گاهی حتی وصول بدهی‌های غیرمالیاتی دولت هم به عهده سازمان‌های مالیاتی گذاشته می‌شود. روش‌های متعددی برای وصول بدهی‌های مالیاتی وجود دارد که یکی از آن‌ها مدیریت بدهی‌های مالیاتی است. مؤدیانی که در این مرحله اقدام به پرداخت مالیات خود نمی‌نمایند، از سطح تمکین پایینی برخوردار بوده و از طرفی هزینه‌های دریافت مالیات برای این دسته از مؤدیان بسیار بالا می‌باشد؛ لذا به‌منظور استفاده بهینه از منابع، انجام مدیریت و برنامه‌ریزی در خصوص بدهی‌های مالیاتی امری ضروری است. مدیریت بدهی‌های مالیاتی به معنی وصول بدهی در کوتاه‌ترین زمان و کم‌ترین هزینه ممکن بوده که استفاده از روش‌های داده‌کاوی به تحقق این اهداف کمک شایانی می‌نماید.

**فرایند خدمات مؤدیان مالیاتی:** این فرایند شامل تمام فعالیت‌هایی است که به مؤدیان در حل سؤالات آنان در زمینه تسلیم اظهارنامه، پرداخت مالیات، حسابرسی و سایر خدمات سازمان‌های مالیاتی کمک می‌نماید. این خدمات می‌تواند از طریق مرکز تماس تلفنی، سایت اینترنتی، پست الکترونیکی و مراجعه به اداره‌های مالیاتی صورت گیرد. با دریافت نظرات مؤدیان مالیاتی در خصوص خدمات گوناگون و ارتقاء آن‌ها، سطح تمکین داوطلبانه مؤدیان افزایش می‌یابد.

### چارچوب داده‌کاوی

داده‌کاوی فرایند کشف دانش و بینش مفید در پایگاه داده‌های بزرگ است. داده‌کاوی یک فناوری است که روش‌های تجزیه و تحلیل سنتی داده را با الگوریتم‌های پیچیده، به‌منظور پردازش حجم زیادی از داده‌ها، ترکیب می‌کند. همچنین فرصت شگفت‌انگیزی برای کاوش و تحلیل انواع جدید داده‌ها ارائه می‌دهد. داده‌کاوی بخشی جدایی‌ناپذیر از فرایند

کشف دانش در پایگاه‌های داده است و در واقع روند کلی تبدیل داده‌های خام به دانش مفید و کاربردی می‌باشد. این فرایند شامل سه فاز اصلی است: پیش‌پردازش داده، داده‌کاوی و پس‌پردازش.

به‌طور کلی وظایف داده‌کاوی به دو دسته اصلی پیش‌بینانه و توصیفی تقسیم می‌شوند. در وظیفه پیش‌بینانه، هدف پیش‌بینی مقدار یک ویژگی براساس مقادیر سایر ویژگی‌ها است. ویژگی که پیش‌بینی می‌شود، معمولاً به‌عنوان متغیر هدف یا وابسته شناخته می‌شود و ویژگی‌های مورد استفاده در انجام پیش‌بینی، متغیرهای توضیحی یا مستقل نامیده می‌شوند. در وظیفه توصیفی، هدف این است که الگوهایی (هم‌بستگی‌ها، روندها، خوشه‌ها و باهم‌آیی‌ها) استخراج شود که به‌طور خلاصه روابط اساسی در داده‌ها را توصیف کنند.

هر یک از وظایف اصلی داده‌کاوی، شامل تکنیک‌هایی است و برای پیاده‌سازی هر یک از تکنیک‌ها، الگوریتم‌های مختلفی وجود دارند. تکنیک‌های حوزه پیش‌بینی شامل رگرسیون، «رده‌بندی و پیش‌بینی» و سری‌های زمانی هستند. تکنیک‌های حوزه توصیفی شامل خوشه‌بندی، تحلیل انجمنی و کاوش دنباله‌ای می‌باشند. در ادامه به شرح هر یک از تکنیک‌ها می‌پردازیم. برخی از تکنیک‌های اصلی مرتبط مانند مصورسازی و تشخیص ناهنجاری نیز در این تحقیق در نظر گرفته شده‌اند که در ادامه به‌صورت خلاصه توضیح داده می‌شود. برای مثال تشخیص ناهنجاری هم با تکنیک‌های توصیفی و هم با روش‌های پیش‌بینانه قابل انجام است.

**خوشه‌بندی:** به‌دنبال یافتن آن دسته از مشاهدات است که با هم نزدیک و مرتبط هستند به‌طوری که مشاهداتی که به یک خوشه تعلق دارند نسبت به مشاهدات متعلق به سایر خوشه‌ها تشابه بیشتری به یکدیگر و تفاوت بیشتری با مشاهدات سایر خوشه‌ها داشته باشند.

**تحلیل انجمنی:** تحلیل انجمنی برای کشف الگوهایی استفاده می‌شود که بیانگر بروز باهم‌آیی آیت‌ها و ویژگی‌ها در داده است. الگوهای کشف‌شده معمولاً در قالب قوانین اگر - آنگاه مبتنی بر زیرمجموعه‌های مشخصی ارائه می‌شوند.

**کاوش دنباله‌ای:** هدف اصلی در کشف الگوهای دنباله‌ای، استخراج توالی‌ها و اقلامی است که در طول زمان به‌صورت مکرر و پی‌درپی اتفاق افتاده‌اند. به عبارت دیگر به‌دنبال استخراج الگوهای دنباله‌ای هستیم که نسبت به سایر الگوهایی از این نوع پرتکرارتر بوده و از یک حداقل قراردادی بیشتر اتفاق افتاده باشند.

**رده‌بندی و پیش‌بینی:** «رده‌بندی و پیش‌بینی» به عمل ساختن مدل برای تشخیص و پیش‌بینی متغیر هدف به‌عنوان تابعی از متغیرهای توضیحی اطلاق می‌شود. لفظ رده‌بندی برای متغیرهای هدف گسسته و عبارت پیش‌بینی برای متغیرهای هدف پیوسته استفاده می‌گردد. برای مثال درخت تصمیم، رده‌بندی مبتنی بر قاعده، شبکه‌های عصبی، ماشین بردار پشتیبان و بیز ساده را می‌توان به‌عنوان روش‌های اصلی «رده‌بندی و پیش‌بینی» نام برد. هر تکنیک از یک الگوریتم یادگیری برای شناسایی مدلی استفاده می‌کند که به بهترین وجه برای ارتباط بین مجموعه ویژگی‌ها و برچسب کلاس داده‌های ورودی مناسب باشد و بتواند برچسب‌های کلاس رکوردهایی که قبلاً ندیده را به درستی پیش‌بینی کند.

**رگرسیون:** رگرسیون یکی از روش‌هایی است با قدمت و محبوب در حوزه پیش‌بینی که برای پیش‌بینی مقادیر داده‌های عددی استفاده می‌شود. رگرسیون به مطالعه وابستگی متغیر وابسته به متغیرهای توضیحی می‌پردازد که با تخمین یا پیش‌بینی مقدار متوسط یا میانگین مقادیر نوع اول در حالتی که مقادیر متغیر نوع دوم معلوم یا معین باشند، صورت می‌گیرد.

**سری‌های زمانی:** سری‌های زمانی مشاهداتی هستند که در طول زمان گردآوری می‌شوند. هر متغیری که در بسیاری از نقاط از زمان مشاهده یا اندازه‌گیری شود، می‌تواند یک سری زمانی تشکیل دهد. مانند قیمت یک سهام خاص یا مقدار یک شاخص کل در بازار بورس، مقدار تقاضا، تولید و یا فروش محصولات یک شرکت. هدف از تحلیل سری‌های زمانی، توصیف رفتار یک سری زمانی از لحاظ تغییرات مقطعی و درازمدت و پیش‌بینی مقادیر آینده است.

**مصورسازی:** مصورسازی ابزاری برای ایجاد بینش در کاربر نسبت به داده است؛ به گونه‌ای که داده خام به تصویری دقیق و مؤثر تبدیل شود تا فهم سریعی را در خصوص داده برای تحلیلگر و کاربر مشخص کند. به عبارت دقیق‌تر، مصورسازی به ارائه بصری اطلاعات برای تعامل بهتر و مؤثرتر تحلیلگر و کاربر با آنچه در داده است، اطلاق می‌شود.

**تشخیص ناهنجاری:** تشخیص ناهنجاری به شناسایی مشاهداتی که ویژگی‌هایشان با بقیه داده‌ها تفاوت قابل توجهی دارد و با عنوان ناهنجاری یا نقطه پرت شناخته می‌شوند، می‌پردازد. هدف الگوریتم‌ها، کشف ناهنجاری‌های واقعی و جلوگیری از برچسب زدن اشیای نرمال به‌عنوان ناهنجار است (تن، استینباچ و کومار<sup>۱</sup>، ۲۰۱۶ و هن، پی و کمبر<sup>۲</sup>، ۲۰۱۱).

### پیشینه پژوهش: کاربردهای داده‌کاوی در فرایندهای مختلف مالیاتی

در این بخش پیشینه پژوهشی کاربردهای داده‌کاوی در فرایندهای مالیاتی، بر اساس چارچوب پیشنهادی ارائه می‌گردد. مقالاتی که مشارکت تئوری و کاربردی بهتری داشته و منابع معتبرتر و مرجعی محسوب می‌شوند، در این بخش ارائه می‌گردند.

#### کاربردهای داده‌کاوی در فرایند ارسال

در پژوهش ساتو، عابدین، خانوم، کونیچه و کیسر<sup>۳</sup> (۲۰۲۱)، روش‌های انتخاب ویژگی و رده‌بندی برای تعیین میزان تمکین مؤدیان کشور فلاند در پرداخت مالیات به کار گرفته شده است. ایشان با استفاده از تکنیک‌های مختلف انتخاب ویژگی و معیار بهره اطلاعات، ابتدا ویژگی‌های اولیه را تعیین نموده و سپس از تکنیک‌های مختلف رده‌بندی مانند درخت تصمیم، پرسپترون چند لایه و شبکه‌های بیزین استفاده نموده‌اند. روش شبکه‌های بیزین بهترین نتایج را داشته است. داسیلوا، کاروالهو و سوزا<sup>۴</sup> (۲۰۱۵) به بررسی درخواست‌های استرداد مالیات پرداخته‌اند. در این پژوهش الگوریتم‌های داده‌کاوی به درک بهتر از مؤدیانی که قوانین را رعایت ننموده و تقاضای استرداد مالیاتی نموده، کمک می‌نماید. با استفاده از مدل‌های پیش‌بینی، مؤدیانی که به احتمال زیاد درخواست استرداد آن‌ها رد می‌گردد، انتخاب می‌شوند. نتایج نشان‌دهنده کارایی این روش می‌باشد.

#### کاربردهای داده‌کاوی در فرایند بازرسی (آزمون)

یک از روش‌های علمی که موجب افزایش کارایی و اثربخشی فرایند بازرسی می‌شود، انجام حسابرسی مالیاتی بر مبنای ریسک می‌باشد. در این نوع از حسابرسی، انتخاب هر مودی برای حسابرسی بر اساس سطح ریسک آن مودی صورت

1. Tan, Steinbach & Kumar
2. Han, Pei & Kamber
3. Satu, Abedin, Khanom, Ouenniche & Kaiser
4. da Silva, Carvalho & Souza

می‌پذیرد. تعاریف متعددی از ریسک مالیاتی در مقالات و منابع علمی ارائه شده که عمدتاً به تفاوت مالیات قطعی و ابرازی مودی اشاره دارد. به‌عنوان نمونه، ساری و اعتمادی (۱۳۹۷) بیان می‌دارند که اصطلاح ریسک به‌عنوان پراکندگی تعریف می‌شود که در میزان عدم اطمینان نسبت به آینده نمود پیدا می‌کند (بریلی و مایرز، ۱۹۹۱). ایشان مطابق نظر نیومن، امر و اسپچمیدت<sup>۱</sup> (۲۰۱۳)، مفهوم کلی ریسک را به ریسک مالیاتی شرکت مرتبط ساخته و بیان می‌دارند که ریسک مالیاتی، به پتانسیلی که یک فعالیت یا عمل، باعث ایجاد تفاوت بین مبلغ مالیات قطعی با مبلغ برآوردی اولیه می‌شود، اشاره دارد. به عبارت دیگر تفاوت مالیات قطعی (تعیینی) با ابرازی شرکت‌ها به واسطه عدم اطمینان نسبت به پرداخت مالیات در آینده، به‌عنوان ریسک مالیاتی و به عبارتی شکاف مالیاتی در نظر گرفته می‌شود (قادری، کفعمی و کریمی حساری، ۱۳۹۷ و ابراهیمی، وطن پرست، رضایی و محمدی نوده، ۱۴۰۰).

تحقیقات متعددی در ارتباط با استفاده از تکنیک‌های داده کاوی در فرایند بازرسی صورت گرفته است. برای نمونه هسو، پاتاک، سرواستاوا، تی‌اسچیدا و بی‌جورکلود<sup>۲</sup> (۲۰۱۵) نحوه پیاده‌سازی روش‌های داده کاوی در اداره مالیاتی ایالت مینوسوتا آمریکا را تشریح نموده و از داده‌های واقعی برای حسابرسی استفاده کرده‌اند. بررسی‌های انجام شده در این تحقیق نشان داده که شکاف (فرار) مالیاتی در سال‌های ۲۰۰۱ و ۲۰۰۶ زیاد بوده و دولت تنها بخشی کوچکی از آن را شناسایی نموده است. موضوع مهم در خصوص کشف فرار مالیاتی، تشخیص مؤدیانی است که دارای شکاف مالیاتی می‌باشند و تکنیک‌های داده کاوی در این راستا به دولت‌ها کمک نموده‌اند. در واقع هدف، کشف تمام مؤدیان کم اظهار نیست و استفاده از روشی برای تخصیص بهینه منابع محدود، مد نظر می‌باشد. روش‌های استفاده شده در این تحقیق درخت تصمیم، بیز ساده، شبکه عصبی پرسپترون چند لایه و ماشین بردار پشتیبان می‌باشند. انتخاب ویژگی بر اساس روش آزمون و خطا و با استفاده از نظر خبرگان انجام شده است. نتایج پیاده‌سازی تحقیق نشان داده که روش‌های داده کاوی، کارایی فرایند انتخاب مودی را به میزان ۶۳ درصد ارتقاء می‌بخشند. همچنین نتایج واقعی به‌دست آمده از حسابرسان بخش مالیاتی نشان داده که رویکرد داده کاوی برای انجام حسابرسی مبتنی بر ریسک مفید می‌باشد. این تحقیق می‌تواند در سایر فرایندهای مالیاتی نیز پیاده‌سازی گردد.

در تحقیق لیو، پن و چن<sup>۳</sup> (۲۰۱۰) تأکید شده که مؤدیان از روش‌های مختلفی برای فرار مالیاتی استفاده می‌کنند و این موضوع موجب بروز مشکل در حسابرسی مالیاتی گردیده است. حسابرسی دقیق و سریع میزبان مالیاتی بسیار با اهمیت می‌باشد. روش‌های سنتی انتخاب برای حسابرسی اصولاً مبتنی بر اطلاعات گذشته یا نظر خبرگان مالیاتی می‌باشد. در این مقاله از روش خوشه‌بندی سلسه مراتبی در انتخاب برای حسابرسی در داده‌های ۳۰ شرکت استفاده شده و در نهایت نتایج با موارد واقعی مورد مقایسه قرار گرفته است. در این مقاله از ۷ شاخص مالی شامل نرخ کلی مالیات، نرخ واقعی مالیات، نرخ سرمایه‌گذاری، نرخ نقدینگی آنی، دارایی به حاشیه سود خالص، نرخ بهای تمام شده به فروش، نرخ هزینه مالی و همچنین ویژگی‌های مالیاتی استفاده شده است. در عمل مشکلات پیش‌بینی نشده زیادی وجود داشته که به‌صورت مستمر مورد بررسی قرار گرفته و نهایتاً موجب ارتقاء مدل گردیده است. در تحقیق روان، یین، دنگ، ژنگ و

<sup>1</sup>. Neuman, Omer & Schmidt

<sup>2</sup>. Hsu, Pathak, Srivastava, Tschida & Bjorklund

<sup>3</sup>. Liu, Pan & Chen

کواين<sup>۱</sup> (۲۰۱۹) در مجموعه شرکت‌های یک گروه، فرار مالیاتی مبتنی بر معاملات اشخاص وابسته (ATTE)، بررسی شده است. انجام معاملات به ظاهر قانونی بین شرکت‌های گروه، روش جدیدی برای فرار مالیاتی بوده و در واقع این معاملات با تنوع و پیچیدگی خود نهایتاً منجر به پرداخت کمتر مالیات می‌گردند.

مالیات یکی از منابع مهم درآمدی دولت‌ها بوده که با استفاده از آن می‌توانند هزینه‌ها و سایر مخارج خود را مدیریت کنند. در این میان، فعالین اقتصادی و شرکت‌های بزرگ همراه با شرکت‌های تابعه خود، سعی در کاهش مالیات بر درآمد خود داشته و تلاش دارند از طرق مختلف برای منافع به‌دست آمده، حفاظتی در برابر سازمان مالیاتی ایجاد نمایند. در عمل دو روش برای پرداخت مالیات کمتر وجود دارد: اجتناب مالیاتی و فرار مالیاتی. هدف هر دو روش، دادن مالیات کمتر است؛ اما در نحوه اجرا و روش‌ها متفاوت می‌باشند. حداقل نمودن مالیات از طریق روش‌های مندرج در قوانین مالیاتی، اجتناب مالیاتی نامیده می‌شود. در مقابل استفاده از روش‌های غیرقانونی مانند عدم گزارش درآمد مالیاتی و گزارش هزینه‌هایی که قانونی نبوده یا عدم پرداخت بدهی مالیاتی، اشاره به فرار مالیاتی دارند. موضوع فرار مالیاتی برای سال‌های متمادی مورد توجه سازمان‌های مالیاتی می‌باشد. برای مثال درآمد مالیاتی از دست رفته کشور چین برابر ۱۰ درصد تولید ناخالص ملی، در سال ۲۰۱۳ می‌باشد. به جهت مقابله با این میزان شکاف مالیاتی، دولت چین اقدامات متعددی انجام داده است مانند: ۱. انتخاب به‌صورت موردی و دستی برای حسابرسی، ۲. انتخاب بر اساس اخبار و شواهد واصله و ۳. انتخاب بر اساس محاسبات کامپیوتری مانند تطبیق متناظر و بررسی شاخص‌های غیرعادی صورت‌های مالی. با اجرای این روش‌ها، روش‌های سنتی فرار مالیاتی مانند استفاده از صورت‌حساب‌های غیرواقعی ارزش افزوده و یا استفاده از دست‌کاری حساب‌ها به‌صورت قابل توجهی کاهش یافته است. در حال حاضر کشور چین سعی در استفاده از روش‌های نوین از جمله حسابرسی مبتنی بر یادگیری ماشین و تحلیل گراف دارد. انواع روش‌های حسابرسی باتوجه به نوع و میزان استفاده از فناوری‌های نوین به شرح ذیل است:

۱. روش‌های سنتی حسابرسی مالیاتی: این روش‌ها شامل دستی، بر اساس اخبار و شواهد و کامپیوتری می‌باشند. دو روش اول زمان‌بر و هزینه‌بر بوده و بیشتر متکی به تجربه مأموران مالیاتی می‌باشند. در مقابل روش کامپیوتری متکی بر تسهیل انتخاب مورد حسابرسی از طریق روش‌هایی مانند تطبیق متناظر صورت‌های مالی و بررسی شاخص‌های غیرعادی صورت‌های مالی می‌باشد. روش‌های سنتی قادر به تحلیل داده‌های عظیم مالیاتی به‌صورت مؤثری نبوده و از ابزارهای مناسبی برای جلوگیری از فرار مالیاتی مبتنی بر معاملات اشخاص وابسته برخوردار نیستند.
۲. روش‌های حسابرسی مبتنی بر یادگیری ماشین: برای حل مشکلات حسابرسی سنتی از تکنیک‌های یادگیری ماشین از جمله تحلیل وابستگی، تحلیل خوشه‌بندی، رده‌بندی و شبیه‌سازی و یادگیری تقویتی استفاده می‌گردد.
۳. روش‌های حسابرسی مبتنی بر گراف: روش تشخیص موارد غیرعادی مبتنی بر گراف به بررسی روابط پیچیده شرکت‌های پرداخت‌کننده مالیات کمک شایانی می‌نماید. این روش در حوزه‌های گوناگون از جمله تشخیص

تقلب در کارت‌های اعتباری، تقلب در بیمه‌های سلامت و سایر موارد کاربرد دارد. اخیراً روش تشخیص موارد غیرعادی مبتنی بر گراف در فرار مالیاتی هم مورد استفاده قرار گرفته است. یکی از مواردی که ریسک مالیاتی را افزایش داده احتمال استفاده از فاکتورهای جعلی می‌باشد. گنزالز و ولاسکوآز<sup>۱</sup> (۲۰۱۳) در تحقیقی برای تشخیص مؤدیان صادر کننده فاکتورهای جعلی از دو روش نظارت‌شده و غیرنظارت‌شده استفاده نموده‌اند. ابتدا از دو روش غیرنظارت‌شده خوشه‌بندی یعنی SOM و Neural gas برای تعیین خوشه مؤدیانی که دارای رفتار مشابه بوده استفاده شده است. در این پژوهش ۵۸۲۱۶۱ مودی در قالب دو مجموعه داده شامل شرکت‌های خیلی کوچک تا کوچک و همچنین شرکت‌های متوسط تا بزرگ در کشور شیلی در سال‌های ۲۰۰۵ تا ۲۰۰۷ مورد بررسی قرار گرفته و از روش‌های نظارت‌شده درخت تصمیم، شبکه عصبی و شبکه بیزین برای رده‌بندی رفتار هر خوشه استفاده شده است. این تحقیق در کشف الگوهای فرار مالیاتی و خلق دانش قابل استفاده در حسابرسی مالیاتی، کمک شایانی نموده است.

همچنین قربانی، لایقی و داوودی (۱۳۹۶) به کمک روش‌های داده‌کاوی اقدام به شناسایی مؤدیان صادرکننده صورت‌حساب جعلی نمودند. مطابق نتایج، روش‌های داده‌کاوی متغیرهایی که در مورد بنگاه‌های اقتصادی کوچک و بزرگ و متوسط جهت حسابرسی باید مد نظر قرار بگیرد را تشخیص داده و مدل شبکه عصبی با صحت ۸۸ درصد بر روی داده‌های اعتبارسنجی و با صحت ۸۹ درصد بر روی داده‌های آزمون موفق به کشف فرار مالیاتی شده است. لونا، پالشیکار، اپتی و بهاتچاریا<sup>۲</sup> (۲۰۱۸) در تحقیقی به شناسایی شرکت‌های صوری با استفاده از حساب‌های بانکی پرداخته و مبتنی بر تراکنش‌های ورودی و خروجی حساب‌های بانکی و تکنیک تشخیص ناهنجاری، سعی نموده لیستی از حساب‌هایی که احتمال می‌رود مربوط به شرکت‌های صوری باشند را تشخیص دهند. به دلیل کمبود داده‌های واقعی، از یک شبیه‌ساز استفاده شده که تشخیص را با دقت و صحت بالایی انجام داده است.

### کاربردهای داده‌کاوی در فرایند وصول

فرایند وصول اشاره به بدهی‌های مالیاتی قطعی شده و پرداخت نشده توسط مؤدیان دارد. در این مرحله با وجود بدهی قطعی، مؤدیان مالیاتی از پرداخت بدهی خود خودداری می‌نمایند. از جمله تحقیقات انجام شده در این خصوص، چن و چنگ<sup>۳</sup> (۲۰۱۰) بوده که به بررسی مالیات گواهینامه‌های رانندگی پرداخته‌اند. در این پژوهش روش قابل اطمینانی برای رده‌بندی پرداخت مالیات در دولت ارائه شده و همچنین ویژگی‌های اصلی که بر پرداخت مالیات گواهینامه‌های رانندگی مؤثر بوده، تجزیه و تحلیل گردیده است. در نهایت قواعد معناداری برای یک سیستم مبتنی بر دانش با استفاده از الگوریتم LEM برای سازمان مالیاتی ایجاد شده است.

پلیسنسیا، هالو و لوجان مورا<sup>۴</sup> (۲۰۲۰) نیز با استفاده از روش‌های داده‌کاوی اقدام به تشخیص موارد عدم پرداخت مالیات اقدام نموده‌اند. با توجه به هزینه‌های فراوان، سازمان‌های مالیاتی به دنبال وصول بدهی مالیاتی در سریع‌ترین زمان

1. González & Velásquez
2. Luna, Palshikar, Apte & Bhattacharya
3. Chen & Cheng
4. Placencia, Hallo & Luján-Mora

و در مراحل اولیه فرایندهای مالیاتی می‌باشند. در این تحقیق از روش یادگیری عمیق برای پیش‌بینی بدهی‌های نکول شده استفاده شده است.

### کاربردهای داده‌کاوی در فرایند خدمات مؤدیان مالیاتی

باتس و لهمن<sup>۱</sup> (۲۰۰۳) به اندازه‌گیری میزان ارزش افزوده ایجاد شده توسط داده‌کاوی در سازمان‌ها پرداخته‌اند. چارچوب ارائه شده شامل چگونگی توسعه و ارزیابی روش‌های داده‌کاوی، بررسی قابلیت کاربرد نتایج حاصل از داده‌کاوی و ارزیابی میزان ارزش افزوده آن با استفاده از بوت استریپینگ، می‌باشد و در اداره درآمد داخلی آلمان به اجرا درآمده است. استفاده روزافزون از سایت‌های اینترنتی و شبکه‌های اجتماعی موجب ایجاد منابع عظیمی از داده‌های الکترونیکی گردیده که منبع خوبی برای ورودی تکنیک‌های داده‌کاوی محسوب می‌شود. از این داده‌ها می‌توان به‌منظور تحلیل نظرات کاربران فضای مجازی در زمینه مالیات استفاده کرد. تحقیقی در خصوص متن‌کاوی نظرات مالیاتی شبکه‌های اجتماعی فیس‌بوک و توییتر توسط یوتامی و لوتفی<sup>۲</sup> (۲۰۱۸) در کشور اندونزی انجام شده است. با توجه به ضرورت افزایش تمکین مالیاتی برای افزایش وصول مالیات در کشور اندونزی، کسب نظر آحاد جامعه در خصوص رضایت‌مندی از ارائه خدمات سازمان مالیاتی بسیار مهم می‌باشد. در این پژوهش، انتخاب ویژگی و روش ماشین بردار پشتیبان و معیارهای ارزیابی صحت و F استفاده شده‌اند.

### یافته‌های پژوهش

#### دسته‌بندی تحقیقات انجام شده بر اساس چارچوب ارائه شده

نگاشت و دسته‌بندی ۵۵ مقاله مورد بررسی بر اساس چارچوب پیشنهادی در جدول ۱ با جزئیات کامل شامل فرایند مالیاتی، حوزه کاربردی، تکنیک داده‌کاوی و الگوریتم، آمده است. به‌منظور خلاصه‌سازی، از عنوان مختصر الگوریتم‌ها به شرح ذیل استفاده شده است: ماشین بردار پشتیبان (SVM)، بیز ساده (NB)، شبکه‌های بیزین (BN)، پرسپترون چندلایه (MLP)، درخت تصمیم (DT)، بردار پشتیبان (SV)، جنگل تصادفی (RF)، تست آماری Tukey، شبکه عصبی عمیق (DNN)، یادگیری انتقالی (TR)، یادگیری عمیق (DL)، شبکه عصبی مصنوعی (ANN)، رگرسیون لجستیک (LogR)، فاکتور ناهنجاری محلی (LOF)، روش خط مبنا (BL)، K-نزدیک‌ترین همسایگی (KNN)، شبکه عصبی فازی (FNN)، الگوریتم بهینه‌سازی ازدحام ذرات (PSO)، نقشه‌های خودسازمان‌ده (SOM)، توزیع نرمال چند متغیره (MND)، الگوریتم درخت تصمیم (CART (CART)، رگرسیون خطی (LR)، تحلیل مؤلفه‌های اصلی (PCA)، تجزیه مقادیر ویژه (SVD)، کاوش قاعده (RM)، خوشه‌بندی سلسه‌مراتبی (HC)، درخت تصمیم ID3 (ID3)، درخت تصمیم (C5.0 (C5.0، رگرسیون چندمتغیره (MVR)، یادگیری بدون برچسب و با برچسب مثبت (PUL)، الگوریتم رتبه اعتماد (TRA). همچنین «انتخاب مبتنی بر ریسک برای حسابرسی مالیاتی» و بازرسی (آزمون) نیز به ترتیب با عنوان خلاصه «حسابرسی مبتنی بر ریسک» و بازرسی آمده است. در ادامه تحلیل‌های مختلفی از منظر توزیع فراوانی

1. Bots & Lohman  
2. Utami & Luthfi

فرایندهای مالیاتی و روش‌های داده کاوی و همچنین نشریات مختلفی که این مقالات را چاپ نموده‌اند و روند رشد مقالات ارائه می‌گردد.

### توزیع تحقیقات بر اساس فرایندهای مالیاتی و روش‌های داده کاوی

توزیع تحقیقات انجام شده بر اساس فرایندهای مالیاتی، حوزه‌های کاربردی و روش‌های داده کاوی در جدول ۲ نشان داده شده است. ستون «تعداد» در هر ردیف مربوط به فراوانی مقالاتی است که از تکنیک مربوطه در فرایند مالیاتی و حوزه کاربردی ذکر شده، استفاده کرده‌اند. همان‌طور که مشخص است، تعدادی از پژوهش‌ها از بیش از یک تکنیک استفاده کرده‌اند.

تحلیل انجام شده بر اساس جدول ۲ نشان می‌دهد از میان مقالات مورد مطالعه، ۵۰ مقاله به فرایند «بازرسی (آزمون)» پرداخته‌اند که ۹۰/۱ درصد کل مقالات را به خود اختصاص داده است. از این ۵۰ مقاله، ۹۴ درصد مقالات به حوزه «انتخاب مبتنی بر ریسک برای حسابرسی مالیاتی» پرداخته‌اند. ۴ درصد و ۲ درصد موارد به ترتیب در حوزه‌های کاربردی «شناسایی مؤدیان کد فروش (صوری)» و «شناسایی مؤدیان فاقد اظهارنامه» قرار دارند. بنابراین حوزه کاربردی «انتخاب برای حسابرسی مبتنی بر ریسک» بیشترین مقالات را به خود اختصاص داده و از سوی محققان بیشتر مورد توجه قرار گرفته است. دلیل این امر، اهمیت بالای این حوزه در حداقل نمودن شکاف مالیاتی و افزایش درآمد مالیاتی دولت‌ها و همچنین کاهش هزینه‌های وصول مالیات است. این کاهش از طریق انتخاب برای حسابرسی و به تبع آن حداقل نمودن فرایند «بازرسی (آزمون)» امکان‌پذیر می‌گردد. دسترسی آسان‌تر به داده‌های مورد نیاز در این حوزه کاربردی، نیز تأثیرگذار بوده و امکان انجام تحقیقات بیشتری در این حوزه را فراهم نموده است.

جدول ۱. نگاشت و دسته‌بندی تحقیقات بر اساس چارچوب ارائه شده

ردیف	مرجع	فرایند مالیاتی	حوزه کاربردی	تکنیک	الگوریتم
۱	دینگ، ژانگ، ژای و لی <sup>۱</sup> (۲۰۲۱)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	SVM
۲	ساتو و همکاران (۲۰۲۱)	ارسال	پرداخت مالیات	رده‌بندی و پیش‌بینی	DT و MLP، BN، NB
۳	سیلوا و همکاران <sup>۲</sup> (۲۰۲۱)	بازرسی	حسابرسی مبتنی بر ریسک	تشخیص ناهنجاری	خوشه‌بندی مبتنی بر SV
۴	گنزالز مارتل، هرناندز و منریک دلارا پنیث <sup>۳</sup> (۲۰۲۱)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	RF
۵	ساویک، آتاناسیویچ، ژاکوتویک و کرجیک <sup>۴</sup> (۲۰۲۱)	بازرسی	حسابرسی مبتنی بر ریسک	تشخیص ناهنجاری	الگوریتم خوشه‌بندی HUNOD
۶	ماتوس و همکاران <sup>۵</sup> (۲۰۲۰)	بازرسی	حسابرسی مبتنی بر ریسک	تحلیل انجمنی	Apriori

1. Ding, Zhang, Zhai & Li

2. Silva et al.

3. González-Martel, Hernández & Manrique-de-Lara-Peñate

4. Savić, Atanasijević, Jakovetić & Krejčić

5. Matos et al.



ردیف	مرجع	فرایند مالیاتی	حوزه کاربردی	تکنیک	الگوریتم
۷	جانسون و مورتی <sup>۱</sup> (۲۰۲۰)	بازرسی	حسابرسی مبتنی بر ریسک	تشخیص ناهنجاری	Tukey و تست آماری SVM
۸	کلینتوس و چتریس <sup>۲</sup> (۲۰۲۰)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	یادگیری نیمه‌نظارتی با استفاده از DNN
۹	ژانگ، شی، دونگ، ژنگ و جی <sup>۳</sup> (۲۰۲۰)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	DL و TR، PUL
۱۰	ایپولیتو و لوزانو <sup>۴</sup> (۲۰۲۰)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی و رگرسیون	LogR، DT، RF، ANN، NB و
۱۱	مهتا و همکاران (۲۰۲۰)	بازرسی	حسابرسی مبتنی بر ریسک	خوشه‌بندی و تحلیل لینک	K-means خوشه‌بندی گراف، TRA و
۱۲	دیدیمو و همکاران <sup>۵</sup> (۲۰۲۰)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی و رگرسیون	LogR و RF، MLP، SVM
۱۳	می، دنگ، شی و ژنگ <sup>۶</sup> (۲۰۲۰)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	MLP و PUL
۱۴	ونهوایولد، مارتنز و پیترز <sup>۷</sup> (۲۰۲۰)	بازرسی	حسابرسی مبتنی بر ریسک	تشخیص ناهنجاری	خوشه‌بندی با عرض ثابت، BL و LOF
۱۵	پلیسنسیا و همکاران (۲۰۲۰)	وصول	مدیریت بدهی مالیاتی	رده‌بندی و پیش‌بینی	DNN
۱۶	جی‌هال، کوانیسر، اردچیر و آزوایزی <sup>۸</sup> (۲۰۱۹)	بازرسی	حسابرسی مبتنی بر ریسک	تشخیص ناهنجاری	روش خوشه‌بندی و الگوریتم K-means
۱۷	وو و همکاران <sup>۹</sup> (۲۰۱۹)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	pseudo، RF، PUL و LightGBM
۱۸	یینگ، اکسیا‌کسین و ویجی <sup>۱۰</sup> (۲۰۱۹)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	BN
۱۹	وی و همکاران <sup>۱۱</sup> (۲۰۱۹)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	NN و TR
۲۰	مهتا و همکاران <sup>۱۲</sup> (۲۰۱۹)	بازرسی	حسابرسی مبتنی بر ریسک	رگرسیون و خوشه‌بندی	LogR و خوشه‌بندی طیفی مبتنی بر گراف
۲۱	پرز‌لپز، دلگادو ردریگوز و سانتوس <sup>۱۳</sup> (۲۰۱۹)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	MLP

1. Johnson & Murty
2. Kleanthous & Chatzis
3. Zhang, Shi, Dong, Zheng & Ji
4. Ippolito & Lozano
5. Didimo
6. Mi, Dong, Shi & Zheng
7. Vanhoeyveld, Martens & Peeters
8. Jihal, Ounacer, Ardchir & Azouazi
9. Wu
10. Ying, Xiaoxin & Weige
11. Wei
12. Mehta
13. Pérez López, Delgado Rodríguez, & de Lucas Santos

ردیف	مرجع	فرایند مالیاتی	حوزه کاربردی	تکنیک	الگوریتم
۲۲	لونا و همکاران (۲۰۱۸)	بازرسی	شناسایی مؤدیان کد فروش	تشخیص ناهنجاری	LOF و FastVOA، RRS
۲۳	مابی - مادیسا <sup>۱</sup> (۲۰۱۸)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	رویکردهای گروهی مبتنی بر NB و DT
۲۴	جوپری و سارنو <sup>۲</sup> (۲۰۱۸)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	DT، SVM، KNN، NB و MLP
۲۵	دی رواکس و همکاران <sup>۳</sup> (۲۰۱۸)	بازرسی	حسابرسی مبتنی بر ریسک	تشخیص ناهنجاری	روش‌های خوشه‌بندی طیفی مبتنی بر گراف
۲۶	ژو، یو، روان، ژنگ و دنگ <sup>۴</sup> (۲۰۱۸)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	SVM، MLP، Adaboost و LightGBM
۲۷	اکسیانجیو، یولین و کیچنگ <sup>۵</sup> (۲۰۱۸)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	ANN
۲۸	رحیمی کیا، محمدی، رحمانی و غضنفری (۲۰۱۷)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی و رگرسیون	MLP، SVM و LogR
۲۹	سمعی راد و شاه‌بهرامی (۲۰۱۶)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی و رگرسیون	SVM و LR
۳۰	ونگ، یو و لی (۲۰۱۶)	بازرسی	شناسایی مؤدیان فاقد اظهارنامه	رده‌بندی و پیش‌بینی	FNN، PSO و Adaboost
۳۱	اکر، اُشیریه‌پر و جان - اُتومو <sup>۶</sup> (۲۰۱۶)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	ANN و DT
۳۲	اسیلگو و همکاران <sup>۷</sup> (۲۰۱۶)	بازرسی	حسابرسی مبتنی بر ریسک	تشخیص ناهنجاری	SOM و توزیع MVN
۳۳	داسیلوا و همکاران (۲۰۱۵)	ارسال	استرداد مالیات	رده‌بندی و پیش‌بینی و رگرسیون	NB، RF و LogR
۳۴	چی‌ها، سینق و چادهوری <sup>۸</sup> (۲۰۱۵)	بازرسی	شناسایی مؤدیان فاقد اظهارنامه	خوشه‌بندی و رده‌بندی و پیش‌بینی	K-means و CART
۳۵	گالمر و لابرو <sup>۹</sup> (۲۰۱۵)	بازرسی	حسابرسی مبتنی بر ریسک	رگرسیون	LR
۳۶	ماتوس، دی‌ماسدو و مونتیرو <sup>۱۰</sup> (۲۰۱۵)	بازرسی	حسابرسی مبتنی بر ریسک	تحلیل انجمنی	PCA، Apriori و SVD
۳۷	هسو و همکاران (۲۰۱۵)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	DT، NB، MLP و SVM

1. Mabe-Madisa

2. Jupri &amp; Sarno

3. De Roux, Perez, Moreno, Villamil &amp; Figueroa

4. Zhu, Yan, Ruan, Zheng &amp; Dong

5. Xiangyu, Youlin &amp; Qicheng

6. Okoro, Oshoiribhor &amp; John-Otumu

7. Assylbekov et al.

8. Jha, Singh &amp; Chaudhary

9. Gallemore &amp; Labro

10. Matos, de Macedo &amp; Monteiro

ردیف	مرجع	فرایند مالیاتی	حوزه کاربردی	تکنیک	الگوریتم
۳۸	جونکیودی فورچون و همکاران <sup>۱</sup> (۲۰۱۴)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	SVM و NB
۳۹	گنزالز و ولاسکوایز (۲۰۱۳)	بازرسی	حسابرسی مبتنی بر ریسک	خوشه‌بندی و رده‌بندی و پیش‌بینی	DT, Neural gas, SOM, MLP و BN
۴۰	امیور و تی کیوات <sup>۲</sup> (۲۰۱۲)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	DT
۴۱	لین و همکاران (۲۰۱۲)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	ANN و DT
۴۲	لیو، شو، شو و ژنگ <sup>۳</sup> (۲۰۱۲)	بازرسی	حسابرسی مبتنی بر ریسک	خوشه‌بندی	K-means
۴۳	وو و همکاران (۲۰۱۲)	بازرسی	حسابرسی مبتنی بر ریسک	تحلیل انجمنی	Apriori
۴۴	یو، چن، گیواینگ، ما و لی‌یو (۲۰۱۱)	بازرسی	حسابرسی مبتنی بر ریسک	خوشه‌بندی	C-Means
۴۵	کالیو و بک (۲۰۱۱)	بازرسی	حسابرسی مبتنی بر ریسک	خوشه‌بندی	SOM
۴۶	چن، سو، چنی هسو و وو <sup>۴</sup> (۲۰۱۱)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	HRCNN و LVQ, MLPs و DT
۴۷	ونگ (۲۰۱۰)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	Rough Sets و RM
۴۸	لی‌یو، پن و چن (۲۰۱۰)	بازرسی	حسابرسی مبتنی بر ریسک	خوشه‌بندی	HC
۴۹	چن و چنگ (۲۰۱۰)	وصول	مدیریت بدهی مالیاتی	رده‌بندی و پیش‌بینی	DT, Rough set و دلفی
۵۰	ونگ و شی (۲۰۰۹)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	ANN
۵۱	ژو و همکاران (۲۰۰۹)	بازرسی	حسابرسی مبتنی بر ریسک	تحلیل انجمنی	Apriori
۵۲	ویلیامز و کریستن <sup>۵</sup> (۲۰۰۷)	بازرسی	حسابرسی مبتنی بر ریسک	خوشه‌بندی	SOM
۵۳	دبار و ایلر - واکر (۲۰۰۶)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	SVM
۵۴	یو، کوین و جای <sup>۶</sup> (۲۰۰۳)	بازرسی	حسابرسی مبتنی بر ریسک	رده‌بندی و پیش‌بینی	DT
۵۵	یوتامی و لوفتی (۲۰۱۸)	خدمات مؤدیان	تحلیل نظرات مالیاتی	رده‌بندی و پیش‌بینی	SVM

1. Junqué de Fortuny et al.

2. Ameer &amp; Tkiouat

3. Liu, Xu, Xu &amp; Zhang

4. Chen, Su, Chen, Hsu &amp; Wu

5. Williams &amp; Christen

6. Yu, Qin &amp; Jia

از میان ۵ مقاله باقی‌مانده، ۲ مورد به فرایند «ارسال» و حوزه «پرداخت مالیات» و «استرداد مالیات» و ۲ مقاله به فرایند «وصول» و حوزه «مدیریت بدهی مالیاتی» اختصاص یافته‌اند. با توجه به کاربردپذیری تکنیک‌های داده‌کاوی در این دو فرایند و حوزه‌های مربوطه و تحقیقات اندکی که صورت گرفته، این زمینه‌ها می‌تواند موضوع مناسبی جهت تحقیقات آتی به حساب آید. در نهایت، تنها یک مقاله به فرایند «خدمات مؤدیان مالیاتی» و حوزه «تحلیل نظرات مالیاتی» پرداخته است. تحلیل نظرات کاربران با استفاده از تکنیک‌های داده‌کاوی در سایت‌های مرتبط با سازمان مالیاتی و سایت‌های خبری می‌تواند مبنای خوبی برای تحقیقات آتی باشد.

محبوب‌ترین و پرکاربردترین تکنیک، «رده‌بندی و پیش‌بینی» است که شامل ۳۵ مقاله می‌شود. تکنیک‌های خوشه‌بندی، تشخیص ناهنجاری، رگرسیون، تحلیل انجمنی و مصورسازی به ترتیب با فراوانی ۸، ۷، ۴ و ۱ در رتبه‌های بعدی قرار دارند.

جدول ۲. توزیع تحقیقات بر اساس فرایند مالیاتی، حوزه کاربردی و تکنیک داده‌کاوی

تعداد	منابع (ردیف مقالات در جدول ۱)	تکنیک داده‌کاوی	حوزه کاربردی	فرایند مالیاتی
۲	۳۳-۲	رده‌بندی و پیش‌بینی	پرداخت مالیات	ارسال
۱	۳۳	رگرسیون	استرداد مالیات	
۱	۲۲	تشخیص ناهنجاری	شناسایی مؤدیان کد فروش (صوری)	بازرسی (آزمون)
۱	۳۴	خوشه‌بندی	شناسایی مؤدیان فاقد اظهارنامه	
۲	۳۴-۳۰	رده‌بندی و پیش‌بینی		
۷	۵۲-۴۸-۴۵-۴۴-۴۲-۳۹-۱۱	خوشه‌بندی		
۴	۵۱-۴۳-۳۶-۶	تحلیل انجمنی		
۲۸	-۲۱-۱۹-۱۸-۱۷-۱۳-۱۲-۱۰-۹-۸-۴-۱ -۳۸-۳۷-۳۱-۲۹-۲۸-۲۷-۲۶-۲۴-۲۳ ۵۴-۵۳-۵۰-۴۷-۴۶-۴۱-۴۰-۳۹	رده‌بندی و پیش‌بینی	انتخاب مبتنی بر ریسک برای حسابرسی مالیاتی	
۶	۳۵-۲۹-۲۸-۲۰-۱۲-۱۰	رگرسیون		
۷	۳۲-۲۵-۱۶-۱۴-۷-۵-۳	تشخیص ناهنجاری		
۱	۱۲	مصورسازی		
۲	۴۹-۱۵	رده‌بندی و پیش‌بینی	مدیریت بدهی مالیاتی	وصول
۱	۵۵	رده‌بندی و پیش‌بینی	تحلیل نظرات مالیاتی	خدمات مؤدیان مالیاتی

جدول ۳ فراوانی کاربرد هر یک از تکنیک‌های داده‌کاوی به تفکیک فرایندها و جدول ۴، تعداد کاربرد هر یک از الگوریتم‌ها را نشان می‌دهد. نتایج بیانگر این موضوع است که به ترتیب الگوریتم‌های شبکه عصبی، درخت تصمیم و ماشین بردار پشتیبان، بیشترین کاربرد را داشته‌اند.

## جدول ۳. توزیع فراوانی تکنیک‌های مختلف در فرایندهای مالیاتی

تکنیک	ارسال	بازرسی (آزمون)	خدمات مؤدیان	وصول
رده‌بندی و پیش‌بینی	۲	۳۰	۱	۲
تحلیل انجمنی	۰	۴	۰	۰
تشخیص ناهنجاری	۰	۸	۰	۰
خوشه‌بندی	۰	۸	۰	۰
رگرسیون	۱	۶	۰	۰
مصورسازی	۰	۱	۰	۰

## جدول ۴. توزیع فراوانی الگوریتم‌های داده‌کاوی مورد استفاده در تحقیقات

ردیف	الگوریتم داده‌کاوی	تعداد	ردیف	الگوریتم داده‌کاوی	تعداد
۱	شبکه عصبی	۱۹	۱۴	LightGBM	۲
۲	درخت تصمیم	۱۳	۱۵	Rough Sets	۲
۳	ماشین بردار پشتیبان	۱۲	۱۶	تحلیل مولفه‌های اصلی (PCA)	۲
۴	بیز ساده	۷	۱۷	فاکتور ناهنجاری محلی (LOF)	۲
۵	جنگل تصادفی	۵	۱۸	روش pseudo	۱
۶	رگرسیون لجستیک	۵	۱۹	FastVOA	۱
۷	Apriori	۴	۲۰	K- نزدیک‌ترین همسایگی	۱
۸	C-Means و K-means	۵	۲۱	RRS	۱
۹	نقشه‌های خودسازمان ده	۴	۲۲	SVD	۱
۱۰	شبکه‌های یادگیری عمیق	۲	۲۳	خوشه‌بندی HUNOD	۱
۱۱	خوشه‌بندی مبتنی بر گراف	۳	۲۴	خوشه‌بندی سلسه‌مراتبی	۱
۱۲	شبکه‌های بیزین	۳	۲۵	روش‌های خوشه‌بندی با عرض ثابت	۱
۱۳	Adaboost	۲	۲۶	روش خط مبنا (BL)	۱

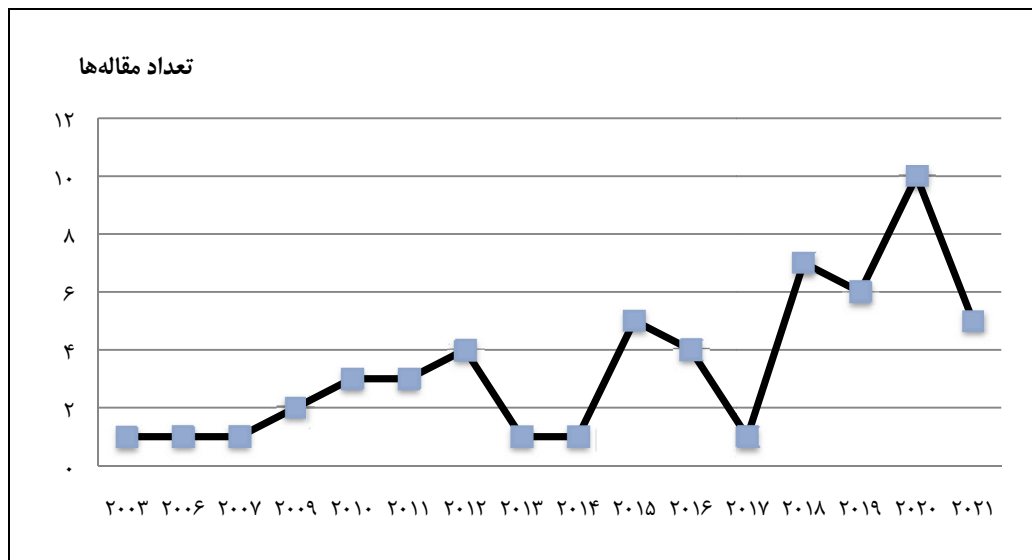
## توزیع تحقیقات بر اساس سال انتشار

مقالات جست‌وجو شده شامل مواردی است که در بازه زمانی ۲۰۰۰ تا ۲۰۲۱ انتشار یافته‌اند. توزیع این مقالات بر اساس سال انتشار در شکل ۲ نمایش داده شده است. روشن است که فراوانی تحقیقات انجام شده در طول زمان به صورت کلی روند روبه‌رشد و صعودی داشته است؛ البته در مقطعی از زمان تعداد مقالات نسبت به سال قبل کاهش یافته است. همچنین در سال ۲۰۲۱ نسبت به سال ۲۰۲۰، یک کاهش قابل توجه در نمودار مشاهده می‌شود که به احتمال زیاد به این دلیل است که مقالات سال ۲۰۲۱ هنوز به صورت کامل چاپ نشده‌اند.

## توزیع تحقیقات بر اساس نشریات

توزیع فراوانی مقالات در نشریات بین‌المللی در جدول ۵ آمده است. مقالات کنفرانسی بین‌المللی نیز در ردیف ۱ آمده است. نشریه Expert Systems with Applications بیشترین فراوانی را به خود اختصاص داده که نشریه شناخته شده

و بسیار معتبری در زمینه سیستم‌های هوشمند، سیستم‌های خبره و کاربردهای داده‌کاوی در کسب و کارهای مختلف می‌باشد.



شکل ۲. توزیع تحقیقات بر اساس سال انتشار

جدول ۵. توزیع فروانی تحقیقات در نشریات

ردیف	عنوان نشریه	تعداد	درصد
۱	Confronce Paper	۳۲	۵۸
۲	Expert Systems with Applications	۴	۷
۳	PAIDEUMA JOURNAL OF RESEARCH	۱	۲
۴	Risk Management	۱	۲
۵	African Journal of Science, Technology, Innovation and Development	۱	۲
۶	Applied Mathematics	۱	۲
۷	Applied Soft Computing	۱	۲
۸	AUTOMATION IN CONSTRUCTION	۱	۲
۹	Decision Support Systems	۱	۲
۱۰	IEEE Access	۱	۲
۱۱	International Journal of Accounting Information Systems	۱	۲
۱۲	International Journal of Artificial Intelligence and Applications (IJAIA)	۱	۲
۱۳	International Journal of Computer Applications	۱	۲
۱۴	Journal of Accounting and Economics	۱	۲
۱۵	Journal of the Chinese Institute of Engineers	۱	۲
۱۶	Knowledge-Based Systems	۱	۲

## نتیجه‌گیری و پیشنهادها

در حال حاضر ادارات امور مالیاتی با پایگاه‌داده‌های عظیمی مواجهند که تحلیل آن‌ها به شیوه‌های سنتی، به دلیل منابع محدود سازمان و همچنین حجم انبوه داده‌های موجود، امکان‌پذیر نیست. داده‌کاوی می‌تواند عملکرد اثربخشی در بهبود فرایندهای مختلف مالیاتی داشته باشد و در راستای تصمیم‌گیری و اتخاذ رویکردهای مناسب، مؤثر واقع شود. از این‌رو داده‌کاوی و تحلیل داده‌های مالیاتی مورد توجه زیادی از محققان و مدیران اداره‌های مالیاتی قرار گرفته است. در پژوهش حاضر، به مرور ادبیات موجود در زمینه کاربرد تکنیک‌های داده‌کاوی در فرایندهای مالیاتی پرداخته شد. هدف ارائه خلاصه‌ای از پژوهش‌های انجام شده در زمینه کاربرد تکنیک‌های داده‌کاوی جهت بهبود فرایندهای مالیاتی است.

در این راستا بر اساس مطالعه مقالات و با تأیید خبرگان حوزه هوشمندسازی مالیات، یک چارچوب برای دسته‌بندی تحقیقات ارائه شد. این چارچوب در سطح اول بر اساس فرایندهای مالیاتی و در سطح دوم در قالب حوزه‌های کاربردی هر یک از این فرایندها، تعریف شده است. ۵۵ مقاله که بین سال‌های ۲۰۰۰ و ۲۰۲۱ در پایگاه داده‌های معتبر آنلاین چاپ شده بودند، مورد بررسی و تحلیل قرار گرفته و بر اساس این چارچوب دسته‌بندی شدند. انتظار می‌رود نتایج بدست آمده از این گزارش، منجر به ایجاد بینش خوبی برای محققان و علاقه‌مندان دنیای واقعی شود. همچنین سعی شده شکاف‌های تحقیقاتی و موضوعات مناسبی که نیاز به پژوهش‌های بیشتر در آن‌ها وجود دارد، معرفی شوند. از تحلیل‌های انجام شده نتایج ذیل حاصل شد:

تحقیقات قابل توجهی در زمینه کاربردهای داده‌کاوی در مالیات انجام شده و در مجموع در طول زمان روند روبه‌رشد و صعودی داشته است. این تحقیقات کاربردپذیری و پتانسیل‌های کاربردی تکنیک‌های داده‌کاوی در این حوزه را نشان می‌دهد که می‌توانند به‌عنوان ایده‌های کاربردی در سیستم‌های مالیاتی دنیا استفاده شوند.

از میان فرایندهای چهارگانه مالیاتی، فرایند بازرسی و حوزه کاربردی «انتخاب مبتنی بر ریسک برای حسابرسی مالیاتی» حجم بیشتری از تحقیقات را به خود اختصاص داده است. دلیل آن را می‌توان اهمیت بالای این حوزه در حداقل نمودن شکاف مالیاتی، افزایش درآمد مالیاتی دولت‌ها و همچنین کاهش هزینه‌های وصول مالیات دانست؛ چرا که مهم‌ترین و هزینه‌برترین فرایند در سازمان‌های مالیاتی، فرایند بازرسی می‌باشد. این کاهش از طریق انتخاب برای حسابرسی و به تبع آن حداقل نمودن فرایند بازرسی امکان‌پذیر می‌گردد. دسترسی آسان‌تر به داده‌های مورد نیاز نیز بر اقبال محققان به این حوزه تأثیرگذار بوده است.

تعداد اندکی از مقالات شامل ۲ مقاله به فرایند «ارسال» و حوزه‌های «پرداخت مالیات» و «استرداد مالیات» و همچنین ۲ مقاله به فرایند «وصول» و حوزه «مدیریت بدهی مالیاتی» اختصاص یافته‌اند. با توجه به کاربردی بودن تکنیک‌های داده‌کاوی در این دو فرایند و تحقیقات بسیار اندک موجود، این دو مورد می‌تواند موضوع مناسبی جهت تحقیقات آتی به حساب آید.

محبوب‌ترین و پرکاربردترین تکنیک، «رده‌بندی و پیش‌بینی» است و الگوریتم‌های شبکه عصبی، درخت تصمیم و ماشین بردار پشتیبان نیز، بیشترین کاربرد را داشته‌اند.

پیشنهاد می‌شود محققان به ارائه راهکار جهت بهبود سایر فرایندهای مالیاتی که تاکنون کم‌تر مورد توجه قرار

گرفته‌اند از جمله «ارسال»، «وصول» و «خدمات مؤدیان» پردازند. مصاحبه با خبرگان هوشمندسازی مالیات نشان می‌دهد که در تمامی این زمینه‌ها پتانسیل‌های خوبی برای پیاده‌سازی تکنیک‌های داده‌کاوی وجود دارد. در خصوص روش‌های تحلیل داده، رویکردهای یادگیری انتقالی به دلایل محدودیت‌های موجود در داده‌های مالیاتی توصیه می‌شود و جای کار بیشتری در این خصوص وجود دارد. همچنین استفاده از تکنیک‌های یادگیری عمیق و تحلیل کلان داده می‌تواند مورد توجه قرار گیرد. برای استفاده از تکنیک‌های تحلیل گراف نیز می‌توان جایگاه ویژه‌ای در حوزه مالیات تعریف کرد مانند تحلیل و شناسایی معاملات مدور و کشف فرار مالیاتی که می‌تواند اثربخشی ویژه‌ای در بهبود فرایندهای مالیاتی ایجاد کند. ارائه چارچوب‌های کاربردی بومی‌سازی شده برای سیستم‌ها و ادارات امور مالیاتی کشورهای مختلف نیز از خالاهای تحقیقاتی است که جای کار زیادی دارد. یکی از خالاهای تحقیقاتی جدی، یکپارچه‌سازی منابع داده داخلی و خارجی ادارات امور مالیاتی و تحلیل داده‌های حاصل است که می‌تواند اثربخشی بسیار ویژه‌ای در این حوزه ایجاد کند.

## منابع

- ابراهیمی، محسن، وطن‌پرست، محمدرضا، رضایی، فرزین و محمدی نوده، فاضل. (۱۴۰۰). عوامل اثرگذار بر سوگیری حسابرسان مالیاتی در قضاوت‌های حرفه‌ای. *بررسی‌های حسابداری و حسابرسی*، ۲۸ (۲)، ۱۸۱-۲۰۵.
- باباجانی، جعفر و باقری، بهروز. (۱۳۹۶). ارائه مدلی برای اندازه‌گیری تمکین مالیاتی در سطح هر مؤدی مالیاتی. *فصلنامه پژوهشنامه مالیات*، ۳۴ (۸۲)، ۱۱-۴۵.
- قادری، بهمن، کفعمی، مهدی و کریمی حصار، فرشاد (۱۳۹۷). بررسی تأثیر عوامل مالی و غیرمالی مؤثر بر شکاف مالیاتی. *فصلنامه چشم‌انداز حسابداری و مدیریت*، ۱ (۲)، ۱-۱۶.
- مام موسی اگریش، اسرین. (۱۳۹۴). نقش مالیات در توسعه اقتصادی. *دومین کنفرانس بین‌المللی مدیریت، حسابداری و اقتصاد*. شیراز.
- نمازیان، علی، پورحیدری، امید و زینلی، حدیث. (۱۴۰۰). بررسی تأثیر نرخ مؤثر مالیاتی و کیفیت حاکمیت شرکتی بر فرار مالیاتی و فساد مالیاتی. *بررسی‌های حسابداری و حسابرسی*، ۲۸ (۳)، ۵۳۲-۵۰۷.

## References

- Ameur, F., & Tkiouat, M. (2012). Taxpayers fraudulent behavior modeling the use of datamining in fiscal fraud detecting Moroccan case. *Applied Mathematics*, 3 (10), 1207-1213.
- Assylbekov, Z., Melnykov, I., Bekishev, R., Baltabayeva, A., Bissengaliyeva, D., & Mamlin, E. (2016). Detecting value-added tax evasion by business entities of Kazakhstan. *In International Conference on Intelligent Decision Technologies*, pp. 37-49.
- Babajani, J., & Bagheri, B. (2017). Proposing a model for measuring tax compliance at the level of each taxpayer. *Quarterly Journal of Tax Research*, 34 (82), 11-45. (in Persian)



- Baisalbayeva, K., van der Enden, E., Tenan, R., & Flores, R. (2018). *The Data Intelligent Tax Administration: Meeting the challenges of Big Tax Data and Analytics*. PwC & Microsoft.
- Bots, P. W., & Lohman, F. A. (2003). Estimating the added value of data mining: A study for the Dutch Internal Revenue Service. *International Journal of Technology, Policy and Management*, 3 (3-4), 380-395.
- Chen, J. H., Su, M. C., Chen, C. Y., Hsu, F. H., & Wu, C. C. (2011). Application of neural networks for detecting erroneous tax reports from construction companies. *Automation in construction*, 20 (7), 935-939.
- Chen, Y. S., & Cheng, C. H. (2010). A Delphi-based rough sets fusion model for extracting payment rules of vehicle license tax in the government sector. *Expert Systems with Applications*, 37 (3), 2161-2174.
- da Silva, L. S., Carvalho, R. N., & Souza, J. C. F. (2015). Predictive models on tax refund claims-essays of data mining in brazilian tax administration. *In International Conference on Electronic Government and the Information Systems Perspective*, pp. 220-228.
- De Roux, D., Perez, B., Moreno, A., Villamil, M. D. P., & Figueroa, C. (2018). Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach. *In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 215-222.
- DeBarr, D., & Eyler-Walker, Z. (2006). Closing the gap: automated screening of tax returns to identify egregious tax shelters. *ACM SIGKDD Explorations Newsletter*, 8 (1), 11-16.
- Didimo, W., Grilli, L., Liotta, G., Menconi, L., Montecchiani, F., & Pagliuca, D. (2020). Combining network visualization and data mining for tax risk assessment. *IEEE Access*, 8, 16073-16086.
- Ding, N., Zhang, X., Zhai, Y., & Li, C. (2021). Risk assessment of VAT invoice crime levels of companies based on DFPSVM: a case study in China. *Risk Management*, 23 (1), 1-22.
- Dohrmann, T., & Pinshaw, G. (2009). *The road to improved compliance: A McKinsey benchmarking study of tax administrations: 2008-2009*. McKinsey & Company, Washington, DC.
- Ebrahimi, M., Vatanparast, M., Rezaei, F., & Mohammadi Nodeh, F. (2021). Investigating the Factors Affecting the Bias of Tax Auditors in Professional Judgments. *Accounting and Auditing Review*, 28 (2), 181-205. (in Persian)
- Gallemore, J., & Labro, E. (2015). The importance of the internal information environment for tax avoidance. *Journal of Accounting and Economics*, 60 (1), 149-167.
- Ghaderi, B., Kafami, M., & Karimi, F. (2018). The Effect of Financial and Non-Financial Firms Characteristics on Tax Gap. *Journal of Accounting and Management Vision*, 1(2), 1-16. (in Persian)
- González, P. C., & Velásquez, J. D. (2013). Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Systems with Applications*, 40 (5), 1427-1436.
- González-Martel, C., Hernández, J. M., & Manrique-de-Lara-Peñate, C. (2021). Identifying business misreporting in VAT using network analysis. *Decision Support Systems*, 141.

- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Hsu, K. W., Pathak, N., Srivastava, J., Tschida, G., & Bjorklund, E. (2015). Data mining based tax audit selection: a case study of a pilot project at the minnesota department of revenue. In M. Abou-Nasr, S. Lessmann, R. Stahlbock & G. Weiss (Ed.), *Real world data mining applications*, pp. 221-245. Springer, Cham.
- Ippolito, A., & Lozano, A. C. G. (2020). Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of São Paulo. In *Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020)*, pp. 452-459.
- Jha, K., Singh, D., & Chaudhary, D. (2015). An Approach to Determine Non-filer's of Property Tax using Clustering and Classification. *International Journal of Computer Applications*, 111 (11), 15-18.
- Jihal, H., Ounacer, S., Ardchir, S., & Azouazi, M. (2019). Clustering Model of False Positive Elimination in Moroccan Fiscal Fraud Detection. In *International Conference on Advanced Intelligent Systems for Sustainable Development*, pp. 125-130.
- Johnson, S. J., & Murty, M. R. (2020). Detecting High Risk Property Taxpayers Using a New Business Intelligence Model: A Case of New York City Property Tax, *PAIDEUMA JOURNAL OF RESEARCH*, 4 (2), 145-153
- Junqué de Fortuny, E., Stankova, M., Moeyersoms, J., Minnaert, B., Provost, F., & Martens, D. (2014). Corporate residence fraud detection. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1650-1659.
- Jupri, M., & Sarno, R. (2018). Taxpayer compliance classification using C4. 5, SVM, KNN, Naive Bayes and MLP. In *2018 International Conference on Information and Communications Technology (ICOIACT)*, pp. 297-303.
- Kallio, M., & Back, B. (2011). The self-organizing map in selecting companies for tax audit. In *Emerging Themes in Information Systems and Organization Studies*, pp. 347-358.
- Kleanthous, C., & Chatzis, S. (2020). Gated Mixture Variational Autoencoders for Value Added Tax audit case selection. *Knowledge-Based Systems*, 188.
- Lin, C. H., Lin, I. C., Wu, C. H., Yang, Y. C., & Roan, J. (2012). The application of decision tree and artificial neural network to income tax audit: the examples of profit-seeking enterprise income tax and individual income tax in Taiwan. *Journal of the Chinese Institute of Engineers*, 35 (4), 401-411.
- Liu, B., Xu, G., Xu, Q., & Zhang, N. (2012). Outlier detection data mining of tax based on cluster. *Physics Procedia*, 33, 1689-1694.
- Liu, X., Pan, D., & Chen, S. (2010). Application of hierarchical clustering in tax inspection case-selecting. In *2010 International Conference on Computational Intelligence and Software Engineering*, pp. 1-4.
- Luna, D. K., Palshikar, G. K., Apte, M., & Bhattacharya, A. (2018). Finding shell company accounts using anomaly detection. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pp. 167-174.
- Mabe-Madisa, G. V. (2018). A Decision Tree and Naïve Bayes algorithm for income tax prediction. *African Journal of Science, Technology, Innovation and Development*, 10 (4), 401-409.

- Mam Musa Agrifash, A. (1394). The role of taxation in economic development. *In Proceedings of the the Second International Conference on Management, Accounting and Economics*. Shiraz. (in Persian)
- Matos, T., de Macedo, J. A. F., & Monteiro, J. M. (2015). An empirical method for discovering tax fraudsters: A real case study of brazilian fiscal evasion. *In Proceedings of the 19th International Database Engineering & Applications Symposium*, pp. 41-48.
- Matos, T., Macedo, J. A., Lettich, F., Monteiro, J. M., Renso, C., Perego, R., & Nardini, F. M. (2020). Leveraging feature selection to detect potential tax fraudsters. *Expert Systems with Applications*, 145.
- Mehta, P., Mathews, J., Bisht, D., Suryamukhi, K., Kumar, S., & Babu, C. S. (2020). Detecting Tax Evaders Using TrustRank and Spectral Clustering. *In International Conference on Business Information Systems*, pp. 169-183.
- Mehta, P., Mathews, J., Kumar, S., Suryamukhi, K., Babu, C. S., Rao, S. K. V., & Bisht, D. (2019). Big Data Analytics for Tax Administration. *In International Conference on Electronic Government and the Information Systems Perspective*. pp. 47-57.
- Mi, L., Dong, B., Shi, B., & Zheng, Q. (2020). A Tax Evasion Detection Method Based on Positive and Unlabeled Learning with Network Embedding Features. *In International Conference on Neural Information Processing*. pp. 140-151.
- Namazian, A., Pourheidari, O., & Zeinali, H. (2021). Investigating the Effect of Effective Tax Rate and Quality of Corporate Governance on Tax Evasion and Tax Corruption. *Accounting and Auditing Review*, 28 (3), 507-532. (in Persian)
- Neuman, S., Omer, T., & Schmidt, A. (2013). Examining the Association between Tax Risk and Tax Outcomes, SSRN, DOI: 10.2139/ssrn.2215129.
- OECD (2010). *Tax Compliance and Tax Accounting Systems*. TPA (Center for Tax Policy and Administration).
- Okoro, F. M., Oshoiribhor, E. O., & John-Otumu, A. M. (2016). A framework for detecting fraudulent activities in edo state tax collection system using investigative data mining. *International Journal of Artificial Intelligence and Applications (IJAIA)*, 7 (3), 11-21.
- Ordóñez, P. J., & Hallo, M. (2019). Data Mining Techniques Applied in Tax Administrations: A Literature Review. *In 2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG)*, pp. 224-229.
- Pérez López, C., Delgado Rodríguez, M. J., & de Lucas Santos, S. (2019). Tax fraud detection through neural networks: an application using a sample of personal income taxpayers. *Future Internet*, 11 (4), 86. doi:10.3390/fi11040086.
- Pistone P., Roeleveld J., Hattingh J., Nogueira J.F.P., & West C. (2019). *Fundamentals of Taxation*. IBFD, Amsterdam, Netherlands.
- Placencia, J. O., Hallo, M., & Luján-Mora, S. (2020). Detection of Taxpayers with High Probability of Non-payment: An Implementation of a Data Mining Framework. *In 2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1-6.

- Rad, M. S., & Shahbahrami, A. (2016). Detecting high risk taxpayers using data mining techniques. In *2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS)*, pp. 1-5.
- Rahimikia, E., Mohammadi, S., Rahmani, T., & Ghazanfari, M. (2017). Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran. *International Journal of Accounting Information Systems*, 25, 1-17.
- Ruan, J., Yan, Z., Dong, B., Zheng, Q., & Qian, B. (2019). Identifying suspicious groups of affiliated-transaction-based tax evasion in big data. *Information Sciences*, 477, 508-532.
- Satu, M. S., Abedin, M. Z., Khanom, S., Ouenniche, J., & Kaiser, M. S. (2021). Application of Feature Engineering with Classification Techniques to Enhance Corporate Tax Default Detection Performance. In *Proceedings of International Conference on Trends in Computational and Cognitive Engineering*, pp. 53-63.
- Savić, M., Atanasijević, J., Jakovetić, D., & Krejić, N. (2021). Tax Evasion Risk Management Using a Hybrid Unsupervised Outlier Detection Method. arXiv preprint arXiv:2103.01033.
- Silva, J., Solano, D., Fernández, C., Ramos, L. N., Urdanegui, R., Herz, J., & Ovallos-Gazabon, D. (2021). Indicators for Smart Cities: Tax Illicit Analysis Through Data Mining. In *Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*, pp. 929-937.
- Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.
- Utami, E., & Luthfi, E. T. (2018). Text mining based on tax comments as big data analysis using SVM and feature selection. In *2018 International Conference on Information and Communications Technology (ICOIACT)*, pp. 537-542.
- Vanhoeyveld, J., Martens, D., & Peeters, B. (2020). Value-added tax fraud detection with scalable anomaly detection techniques. *Applied Soft Computing*, 86. Doi: org/10.1016/j.asoc.2019.105895
- Wang, J., Yu, X., & Li, P. (2016). Research of tax assessment based on improved Fuzzy Neural Network. In *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, pp. 115-119.
- Wang, X. Z., & Xie, Q. H. (2009). Wavelet neural network model application on the case selection of tax check in real estate industry. In *2009 International Conference on Intelligent Human-Machine Systems and Cybernetics*, Vol. 2, pp. 189-192.
- Wang, Y. (2010). Research on rough sets theory based tax data mining. In *2010 International Conference on Future Information Technology and Management Engineering*, Vol. 1, pp. 13-15.
- Wei, R., Dong, B., Zheng, Q., Zhu, X., Ruan, J., & He, H. (2019). Unsupervised Conditional Adversarial Networks for Tax Evasion Detection. In *2019 IEEE International Conference on Big Data*, pp. 1675-1680.
- Williams, G. J., & Christen, P. (2007). Exploratory multilevel hot spot analysis: Australian taxation office case study. In *Proceedings of the sixth Australasian conference on Data mining and analytics*, Vol. 70, pp. 77-84.

- Wu, R. S., Ou, C. S., Lin, H. Y., Chang, S. I., & Yen, D. C. (2012). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications*, 39 (10), 8769-8777.
- Wu, Y., Zheng, Q., Gao, Y., Dong, B., Wei, R., Zhang, F., & He, H. (2019). TEDM-PU: A Tax Evasion Detection Method Based on Positive and Unlabeled Learning. *In 2019 IEEE International Conference on Big Data*, pp. 1681-1686.
- Xiangyu, X., Youlin, Y., & Qicheng, X. (2018). Intelligent Identification of Corporate Tax Evasion Based on LM Neural Network. *In 2018 37th Chinese Control Conference (CCC)*, pp. 4507-4511.
- Ying, Q., Xiaoxin, H., & Weige, J. (2019). Research on Tax Inspection Case Selection Model Based on Bayesian Network. *In Proceedings of the 2019 2nd International Conference on Information Management and Management Sciences*. pp. 198-202.
- Yu, F., Qin, Z., & Jia, X. L. (2003). Data mining application issues in fraudulent tax declaration detection. *In Proceedings of the 2003 International Conference on Machine Learning and Cybernetics*, 4, 2202-2206.
- Yu, H., Chen, J., Qing, D., Mao, S., & Liu, L. (2011). An application of improved fuzzy c-means clustering algorithm in tax administration, *in IET International Communication Conference on Wireless Mobile and Computing (CCWMC 2011)*, pp. 496-499.
- Zhang, F., Shi, B., Dong, B., Zheng, Q., & Ji, X. (2020). TTED-PU: A Transferable Tax Evasion Detection Method Based on Positive and Unlabeled Learning. *In 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 207-216.
- Zhu, Q. X., Guo, L. J., Liu, J., Xu, N., & Li, W. X. (2009). Research of tax inspection cases-choice based on association rules in data mining. *In 2009 International Conference on Machine Learning and Cybernetics*, Vol. 5, pp. 2625-2628.
- Zhu, X., Yan, Z., Ruan, J., Zheng, Q., & Dong, B. (2018). Irtd-tl: An inter-region tax evasion detection method based on transfer learning. *In 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pp. 1224-1235.