Scientometrics Research Journal
Scientific Bi-Qualterly Journal of Shahed University/ Article in Press (Not edited)

A Co-word Analysis of the Persian Data Mining Articles Indexed in the ISC Database

1. Hamed Barangi[1]
2. Ali Akbar Khasseh*[2]

1. Department of Computer Engineering, Payame Noor University, Tehran, Iran. ORCID: 0000-0003-2444-3474
2. Associate professor, Department of Knowledge and Information Science, Payame Noor University, Tehran, Iran. (Corresponding Author): khasseh@pnu.ac.ir. ORCID: 0000-0001-5664-4671

Abstract

Purpose: The rapid expansion of digital technologies and the increasing availability of large-scale datasets have positioned data mining as one of the most influential research domains. Given its interdisciplinary nature, understanding the co-word structure and thematic development of data-mining research has become increasingly important. Although bibliometric studies have explored international trends using databases such as Scopus and Web of Science, relatively little attention has been devoted to Persian-language research outputs. This study investigates the conceptual structure and thematic evolution of Persian data-mining research indexed in the Islamic World Science Citation Center (ISC) over a twenty-year period (2005–2024). Specifically, it identifies research themes, examines their temporal evolution, maps conceptual relationships among keywords, and determines the strategic positions of themes within the field.

Methodology: This descriptive study has an applied approach and uses bibliometric analysis tools. Data were manually collected from ISC in April 2026 using different variations of the Persian equivalent of the term “data mining.” The initial search retrieved 595 records. After expert review and screening to exclude conceptually irrelevant publications, 589 articles published between 2005 and 2024 were retained for analysis. Because ISC does not provide comprehensive bibliographic exports comparable to those of international databases, data were extracted manually using spreadsheet software. Author keywords were identified and standardized through expert-led data cleaning. Synonyms, singular and plural forms, spelling inconsistencies, and spacing and orthographic variations were merged into unified terms to improve analytical accuracy. The general keyword “data mining” was excluded because of its ubiquitous presence and limited discriminatory value. Biblioshiny was used to generate thematic evolution maps, co-occurrence networks, and strategic diagrams. These analyses enabled the identification of dominant themes, conceptual clusters, emerging topics, and patterns of thematic continuity and transformation over time. They also provided a basis for examining how keyword relationships and theme positions changed across the selected periods.

Findings: The 589 articles contained 1,734 keywords, averaging approximately three keywords per publication. Decision trees (n = 78) and artificial neural networks (n = 74) were the most frequent keywords, followed by clustering (n = 53), prediction (n = 43), and support vector machines (n = 35). Temporal analysis showed the continued presence of established methodological keywords, alongside increased occurrences of “machine learning” and “random forest” in later periods. Because “machine learning” is a broad paradigm, its increased frequency alone does not demonstrate the adoption of a specific advanced algorithm. Thematic evolution analysis showed continuity and branching of selected keywords across the three periods, including links between methodological terms and application domains. “Decision trees” and “clustering” persisted across periods, while some methodological and application-related terms appeared or became more frequent later. For example, thematic flows connected decision trees with support vector machines and artificial neural networks, while clustering was linked with application-related terms such as breast cancer. These patterns describe the observed continuity and branching of keywords and should not be interpreted as evidence of a general transition toward more advanced methods or as a definitive developmental trajectory for the entire field.
Co-word network analysis identified ten conceptual clusters. The network included both methodological keywords and application-related terms, illustrating the coexistence of data-mining methods and the domains in which those methods were used. The clusters therefore capture both algorithmic approaches and their reported application contexts. The interpretation is limited to Persian-language articles indexed in ISC and reflects the co-occurrence structure observed in the selected dataset. Strategic diagram analysis showed that the identified clusters were distributed mainly among the basic, specialized, and emerging/declining quadrants, with no cluster located in the motor-theme quadrant. This distribution should be interpreted cautiously because the network was constructed from keywords occurring at least three times and therefore represents the structure of the available dataset rather than the thematic maturity of data-mining research in Iran as a whole.

Conclusion: The findings suggest a gradual transformation in Persian data-mining research, from an initial focus on foundational algorithms to the application of data-mining techniques in specialized domains. Nevertheless, the absence of mature motor themes and the continued prominence of basic topics indicate that strong, independent thematic trajectories have yet to emerge in the analyzed dataset. The results also suggest that Persian-language research has largely adapted internationally established paradigms, rather than generating distinctive theoretical contributions. Strengthening interdisciplinary collaboration, expanding engagement with emerging research areas, and enhancing connections with international scientific networks may facilitate further thematic development, maturity, and innovation. These findings describe the keyword and co-occurrence patterns observed in the selected Persian-language ISC corpus and should not be generalized to the entire data-mining research output of Iran. Future studies using broader databases, larger keyword networks, and complementary bibliometric analyses could provide a more comprehensive assessment of thematic development and support comparison with international research trends.

Keywords: *Data mining; Bibliometrics; Co-word analysis; Thematic evolution; Islamic World Science Citation Center (ISC).*

Receive:
../../….
Acceptance:
../../….

ویرایش نشده

پژوهش‌نامه علم‌سنجی
دوفصلنامه علمی ـ پژوهشی دانشگاه شاهد/ زودآیند ویرایش نشده

تحلیل هم واژگانی مقالات فارسی حوزه داده‌کاوی نمایه شده در مؤسسه استنادی و پایش علم و فناوری جهان اسلام(آی.اس.سی)

۱. حامد برنگی[۱]
۲. علی‌اکبر خاصه*[۲]

چکیده

هدف: این پژوهش با استفاده از فنون علم سنجی به تحلیل هم واژگانی و تکامل موضوعی پژوهش‌های فارسی حوزه داده‌کاوی در مؤسسه استنادی و پایش علم و فناوری جهان اسلام (آی اس سی) پرداخته است.

روش‌شناسی: پژوهش حاضر از نوع توصیفی با رویکرد کاربردی و مبتنی بر تحلیل علم سنجی است. داده‌های مورد نیاز این پژوهش فروردین ۱۴۰۵ از آی اس سی استخراج گردید. از نرم افزار بیب اکسل برای پردازش اولیه داده‌ها، و از نرم افزار بیبلیوشاینی برای ترسیم نقشه و تحلیل خوشه‌ای داده‌ها استفاده شد.

یافته‌ها: تحلیل ۵۸۹ مقاله فارسی در حوزه داده کاوی، نمایه‌شده در پایگاه آی‌اس‌سی نشان داد که در مجموع ۱۷۳۴ کلیدواژه در این مقالات به‌کار رفته است. «درخت تصمیم» با ۷۸ رخداد و «شبکه عصبی مصنوعی» با ۷۴ رخداد پرتکرارترین کلیدواژه‌ها بودند و پس از آن‌ها «خوشه‌بندی» (۵۳)، «پیش‌بینی» (۴۳) و «ماشین بردار پشتیبان» (۳۵) قرار داشتند. بررسی زمانی کلیدواژه‌ها تداوم حضور برخی مفاهیم روش‌شناختی و افزایش حضور اصطلاحاتی مانند «یادگیری ماشین» و «جنگل تصادفی» را در دوره‌های پایانی نشان داد. با توجه به عام بودن مفهوم «یادگیری ماشین»، افزایش فراوانی آن به‌تنهایی نمی‌تواند بیانگر به‌کارگیری یک الگوریتم خاص یا حرکت به سوی روش‌های پیشرفته‌تر باشد. تحلیل تکامل موضوعی نیز تداوم و انشعاب برخی کلیدواژه‌ها را در سه دوره زمانی نشان داد. همچنین تحلیل هم‌رخدادی به شناسایی ۱۰ خوشه مفهومی انجامید.

نتیجه‌گیری: یافته‌های این پژوهش نشان می‌دهد که پژوهش‌های فارسی حوزه داده‌کاوی در ایران از مرحله آشنایی با روش‌های بنیادین به سوی کاربردی‌سازی در حوزه‌های تخصصی حرکت کرده‌اند، اما هنوز فاصله معناداری با شکل‌گیری مضامین محرک و تولید دانش نظری مستقل دارند. غلبه موضوعات پایه و اتکای نسبی به پارادایم‌های تثبیت‌شده بین‌المللی بیانگر آن است که این حوزه، اگرچه از نظر کاربردی در مسیر رشد قرار گرفته، اما برای دستیابی به بلوغ علمی نیازمند تقویت ظرفیت‌های پژوهشی در زمینه‌های نوظهور، و توسعه همکاری‌های میان‌رشته‌ای است.

واژگان کلیدی: داده‌کاوی، علم‌سنجی، تحلیل هم واژگانی، مؤسسه استنادی و پایش علم و فناوری جهان اسلام (آی‌اس‌سی).

۱. دکتری گروه مهندسی کامپیوتر، دانشگاه پیام نور، تهران، ایران.
۲. دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه پیام‌نور، تهران، ایران (نویسنده مسئول)، khasseh@pnu.ac.ir

دریافت: ۰۰۰/۰۰/۰۰
پذیرش: ۰۰۰/۰۰/۰۰

در آینه ویرایش نشده

مقدمه و بیان مسئله

در عصر تحول دیجیتال، رشد تصاعدی داده‌های دیجیتال در حوزه‌های علمی، صنعتی، اقتصادی و اجتماعی، منجر به شکل‌گیری یکی از مهم‌ترین چالش‌های عصر اطلاعات، یعنی استخراج دانش معنادار از داده‌های عظیم، شده است. توسعه فناوری‌های اطلاعاتی، اینترنت اشیاء، شبکه‌های اجتماعی و سامانه‌های هوشمند، حجم بی‌سابقه‌ای از داده‌های ساخت‌یافته و غیرساخت‌یافته را تولید کرده است که تحلیل آن‌ها بدون بهره‌گیری از روش‌های پیشرفته امکان‌پذیر نیست (Chen, 2017).

داده‌کاوی فرایند استخراج الگوها، روابط و دانش معتبر از مجموعه‌های بزرگ داده است و یکی از مراحل اصلی فرایند کشف دانش در پایگاه‌های داده به شمار می‌رود (Fayyad et al., 1996). این حوزه، با تلفیق مفاهیم آمار، یادگیری ماشین، هوش مصنوعی، پایگاه داده و بازیابی اطلاعات، به یکی از مهم‌ترین زمینه‌های پژوهشی تبدیل شده است (Witten et al., 2011). داده‌کاوی ماهیتی میان‌رشته‌ای دارد و از روش‌ها و تکنیک‌های مختلف برای تحلیل داده‌ها و کشف دانش بهره می‌گیرد؛ ازاین‌رو، امروزه در طیف گسترده‌ای از حوزه‌ها از جمله پزشکی، سلامت، بانکداری، تجارت الکترونیک، مدیریت، آموزش، کشاورزی، حمل‌ونقل و بازارهای مالی به کار گرفته می‌شود (Witten et al., 2011; Sarker, 2021).

رشد سریع پژوهش‌های داده‌کاوی در سال‌های اخیر موجب شده است که تحلیل ساختار علمی، روندهای موضوعی، الگوهای همکاری علمی و تحول مفهومی این حوزه، اهمیت فزاینده‌ای پیدا کند. در این میان، مطالعات علم‌سنجی و کتاب‌سنجی به عنوان ابزارهایی مؤثر برای تحلیل کمی تولیدات علمی، امکان شناسایی روندهای پژوهشی، بازیگران کلیدی، ساختار همکاری‌ها، خوشه‌های موضوعی و موضوعات نوظهور را فراهم می‌کنند (نجار لشگری و همکاران، ۱۴۰۲؛ Donthu et al., 2021).

از سوی دیگر، گسترش کاربردهای داده‌کاوی در ایران نیز موجب افزایش قابل توجه پژوهش‌های مرتبط در مجلات فارسی‌زبان شده است. با این حال، برخلاف پایگاه‌های بین‌المللی نظیر اسکوپوس و وب آو ساینس، مطالعات نظام‌مند اندکی درباره ساختار علمی پژوهش‌های فارسی داده‌کاوی انجام شده است. در این میان، مؤسسه استنادی و پایش علم و فناوری جهان اسلام (آی اس سی) به عنوان مهم‌ترین نظام استنادی کشور، بخش عمده‌ای از تولیدات علمی فارسی را نمایه می‌کند و از این منظر، منبع ارزشمندی برای تحلیل وضعیت پژوهش‌های داخلی به شمار می‌رود. باوجود اهمیت روزافزون داده‌کاوی در فضای علمی ایران، تاکنون مطالعه ای که ساختار مفهومی و تحول موضوعی پژوهش‌های فارسی این حوزه را بررسی کند، مشاهده نشده است.

به طور بالقوه، ماهیت میان‌رشته‌ای داده‌کاوی موجب می شود که پژوهش‌های این حوزه در مجلات و رشته‌های متنوعی پراکنده شوند و دانشگاه‌ها و پژوهشگران حوزه‌های مختلف در آن مشارکت داشته باشند. این پراکندگی موضوعی و سازمانی، ضرورت تحلیل ساختار علمی در این حوزه را دوچندان می‌کند. علم‌سنجی با ابزارهایی چون تحلیل هم‌استنادی، تحلیل هم‌واژگانی و تحلیل شبکه‌ای همکاری‌های علمی، قادر است ساختار و تکامل مفهومی یک حوزه علمی را آشکار سازد (Leydesdorff, 2001; Mokhtarpour & Khasseh, 2020) و با شناسایی موضوعات نوظهور، مضامین پایه، موضوعات محرک و تحلیل روند مفاهیم می‌تواند در سیاست‌گذاری پژوهشی، هدایت تحقیقات آینده و شناسایی خلأهای دانشی نقش مهمی ایفا کند (Cobo et al., 2011).

در این پژوهش، داده‌کاوی حوزه موضوعی عام محسوب می‌شود و مفاهیمی مانند «درخت تصمیم»، «شبکه عصبی مصنوعی»، «ماشین بردار پشتیبان»، «جنگل تصادفی»، «خوشه‌بندی» و «یادگیری ماشین» به‌عنوان روش‌ها، الگوریتم‌ها یا رویکردهای تخصصی به‌کاررفته در پژوهش‌های این حوزه تحلیل شده‌اند؛ همچنین موضوعاتی مانند سلامت، مدیریت ارتباط با مشتری، بانکداری و کشاورزی به‌عنوان حوزه‌های کاربردی داده‌کاوی در نظر گرفته شده‌اند.

بر اساس آنچه گفته شد، پژوهش حاضر تلاش می‌کند از طریق تحلیل هم واژگانی و تحلیل روند موضوعات ، تصویری جامع از وضعیت پژوهش‌های داده‌کاوی در ایران در بازهٔ زمانی سال‌های ۱۳۸۴ تا ۱۴۰۳ ارائه دهد. تحلیل هم‌واژگانی از روش‌های پرکاربرد در کتابسنجی برای ترسیم نقشه‌های دانش است که سال‌هاست به عنوان ابزاری برای تعیین مهم‌ترین موضوع‌های پژوهشی در یک حوزه

با تمرکز بر واژه‌های پربسامد استفاده می‌شود. در این روش که ارتباط میان واژه‌های به‌کاررفته در قسمت‌های مختلف مدارک (از جمله عنوان، چکیده، کلیدواژه‌ها و مانند آن) مورد بررسی قرار می‌گیرد، فرض بر آن است که پربسامدترین واژه‌ها در مقایسه با واژه‌های کم‌بسامد، تأثیر بیشتری بر یک حوزه داشته‌اند. به طور کلی، تحلیل هم‌واژگانی این امکان را فراروی پژوهشگران قرار می‌دهد تا خوشه‌های موضوعی در یک مجله یا حوزه خاص را شناسایی کنند (جواهری و همکاران، ۱۴۰۰؛ خاصه و همکاران، ۱۴۰۰؛ فرهنگی و همکاران، ۱۳۹۷). تحلیل روند موضوعات نیز پویایی زمانیِ کلمات کلیدی را در مجموعهٔ تحت بررسی شناسایی و مصورسازی می‌کند (Aria et al., 2026). این ابزار با ردیابی فراوانی کلمات کلیدی در طول سال‌های انتشار، نشان می‌دهد که کدام موضوعات در حال افزایش هستند، و کدام در حال کاهش هستند و تصویری واضح از تکامل علایق پژوهشی در طول زمان ارائه می‌دهد (Biblioshiny, 2026). با توجه به مطالب فوق، پرسش محوری این است که ساختار مفهومی پژوهش‌های فارسی‌زبان حوزه داده‌کاوی در بازه زمانی ۱۳۸۴ تا ۱۴۰۳ با استفاده از تحلیل هم واژگانی چگونه است؟

سؤال‌های پژوهش

۱. ۱. تکامل موضوعی و مسیر تغییرات موضوعات پژوهشی در حوزه‌ی داده‌کاوی در گذر زمان چگونه بوده است؟

۲. ساختار مفهومی در حوزهٔ داده‌کاوی، بر اساس تحلیل شبکهٔ هم‌رخدادی کلیدواژه‌ها، چگونه است و خوشه‌های موضوعی اصلی در این شبکه کدامند؟

۳. جایگاه و ساختار موضوعات پژوهش‌های فارسی حوزه داده‌کاوی بر اساس نمودار راهبردی چگونه است؟

چارچوب نظری

چارچوب نظری این پژوهش بر تحلیل علم‌سنجی پژوهش‌های حوزه داده‌کاوی استوار است. در این پژوهش، داده‌کاوی موضوع مورد مطالعه است و مفاهیم و روش‌های علم‌سنجی به‌عنوان چارچوب تحلیلی برای بررسی ساختار مفهومی و تکامل موضوعی پژوهش‌های این حوزه به کار گرفته شده‌اند. بر این اساس، تولیدات علمی حوزه داده‌کاوی نه صرفاً مجموعه‌ای از اسناد مستقل، بلکه شبکه‌ای از روابط مفهومی میان موضوعات پژوهشی تلقی می‌شوند.

مبنای نظری داده‌کاوی در چارچوب کشف دانش در پایگاه‌های داده تعریف می‌شود که آن را فرایندی شامل انتخاب داده، پیش‌پردازش، تبدیل، داده‌کاوی و تفسیر نتایج برای استخراج الگوهای معتبر و معنادار از داده‌ها می‌داند (Fayyad et al., 1996). در ادامه، مدل کریسپ ، داده‌کاوی را به‌عنوان یک فرایند چرخه‌ای و مسئله‌محور معرفی می‌کند که میان داده، تحلیل و تصمیم‌گیری پیوند برقرار می‌سازد (Shearer, 2000). این دو چارچوب نشان می‌دهند که داده‌کاوی صرفاً یک ابزار محاسباتی نیست، بلکه یک رویکرد نظام‌مند برای تولید دانش در بسترهای پیچیده داده‌محور به شمار می رود. از آنجا که هدف این پژوهش تحلیل روش‌های داده‌کاوی نیست، بلکه بررسی روند تحول دانش در این حوزه است، چارچوب‌های علم‌سنجی برای مطالعه ساختار مفهومی و تکامل موضوعی پژوهش‌های داده‌کاوی به کار گرفته شده‌اند.

بر اساس نظریه تحول علمی کوهن، توسعه علم در قالب پارادایم‌های مسلط و تغییرات گسسته میان آن‌ها رخ می‌دهد (Kuhn, 1962). در حوزه داده‌کاوی، گذار از روش‌های آماری کلاسیک به رویکردهای یادگیری ماشین، یادگیری عمیق و تحلیل کلان‌داده را می‌توان به‌عنوان یک تحول پارادایمی تفسیر کرد. این تحول بیانگر تغییر در ابزارها، مفاهیم و جهت‌گیری‌های پژوهشی از تحلیل‌های توصیفی به سمت مدل‌های پیش‌بینی و هوشمند است.

در مطالعات علم‌سنجی، «تحول موضوعی» به فرایند تغییر و تداوم مضامین پژوهشی در طول زمان اشاره دارد. این مفهوم نشان می‌دهد که چگونه موضوعات علمی شکل می‌گیرند، توسعه می‌یابند، با یکدیگر ادغام یا از یکدیگر منشعب می‌شوند و در دوره‌های مختلف جایگاه خود را حفظ یا از دست می‌دهند. تحلیل تحول موضوعی معمولاً بر پایه تحلیل هم‌رخدادی کلیدواژه‌ها و مقایسه خوشه‌های مفهومی در بازه‌های زمانی مختلف انجام می‌شود و امکان بررسی پویایی ساختار دانش یک حوزه را فراهم می‌کند (Cobo et al., 2011).

ساختار مفهومی علم در این پژوهش بر مبنای تحلیل هم‌رخدادی واژگان تبیین می‌شود که توسط کالون و همکاران توسعه یافته است. این رویکرد بر این فرض استوار است که هم رخدادی کلیدواژه‌ها در اسناد علمی، نشان‌دهنده ارتباط مفهومی میان آن‌ها و ساختار شناختی یک حوزه علمی است (احمدی و حسین پور، ۱۴۰۴؛ عشقی و همکاران، ۱۴۰۵؛ Callon et al., 1983). بر اساس این دیدگاه، خوشه‌های مفهومی حاصل از تحلیل هم‌رخدادی، بازتاب‌دهنده موضوعات اصلی، نوظهور و تخصصی در یک حوزه علمی هستند. همچنین مدل «نقشه راهبردی» که امکان طبقه‌بندی موضوعات علمی را در چهار دسته «موضوعات محرک»، «موضوعات بنیادی»، «موضوعات تخصصی» و «موضوعات نوظهور یا در حال افول» قرار می‌دهد (سهیلی و همکاران، ۱۳۹۷؛ Khasseh et al., 2017) در این چارچوب، مبنای تحلیل تحول موضوعی در پژوهش حاضر است.

پیشینه پژوهش

در دو دهه اخیر، گسترش فناوری‌های دیجیتال، افزایش حجم داده‌های تولیدشده و پیشرفت‌های چشمگیر در حوزه‌های یادگیری ماشین، کلان‌داده و هوش مصنوعی، موجب رشد سریع پژوهش‌های داده‌کاوی در ابعاد نظری و کاربردی شده است. این توسعه گسترده، علاوه بر افزایش کمّی تولیدات علمی، به شکل‌گیری شاخه‌های میان‌رشته‌ای جدید، تنوع موضوعات پژوهشی و گسترش همکاری‌های علمی منجر شده است. در چنین شرایطی، روش‌های علم‌سنجی و کتاب‌سنجی به‌عنوان ابزارهایی کارآمد برای تحلیل ساختار دانشی، شناسایی روندهای تحول علمی، کشف موضوعات نوظهور و بررسی الگوهای همکاری پژوهشگران مورد توجه قرار گرفته‌اند. ازاین‌رو، مرور و تحلیل مطالعات علم‌سنجی پیشین در حوزه داده‌کاوی و حوزه‌های مرتبط، می‌تواند تصویری روشن از وضعیت موجود پژوهش‌ها ارائه داده و زمینه شناسایی خلأهای پژوهشی و تبیین جایگاه پژوهش حاضر را فراهم سازد.

بوزتاش و همکاران (Boztas et al., 2024) با یک تحلیل کتاب‌سنجی از حوزه داده‌کاوی آموزشی نشان دادند که تعداد انتشارات در این زمینه ۱۳۲۵ درصد رشد کرده و مفاهیمی چون «تحلیل یادگیری» و سپس «یادگیری عمیق» و «عواطف» به‌عنوان مضامین محرک ظهور یافته‌اند. کوزیس و تجورتیس (Kousis & Tjortiis, 2021) نیز با تحلیل کتاب‌سنجی ۱۹۷ مقاله، گستره وسیعی از الگوریتم‌های داده‌کاوی را در لایه‌های مختلف شهر هوشمند شناسایی کردند. در حوزه علوم داده، ورما و همکاران (Verma et al., 2025) با تحلیل بیش از ۴۰ هزار مدرک در اسکوپوس طی سال‌های ۲۰۱۳ تا ۲۰۲۳، گذار از توسعه روش‌شناختی به کاربردهای میان‌رشته‌ای را نشان دادند و یادگیری ماشین، تحلیل پیش‌بینانه و اینترنت اشیاء را به‌عنوان خوشه‌های غالب معرفی کردند. تسنگ و همکاران (Tseng et al., 2016) با مقایسه تطبیقی مقالات داده‌کاوی و کلان‌داده در پایگاه اسکوپوس، کشورها و مجلات اصلی را شناسایی و هم‌پوشانی میان این دو حوزه را نشان دادند.

به طور کلی، بخش قابل توجهی از کارهای کتاب‌سنجی، کاربرد داده‌کاوی را در بخش‌های مختلف ترسیم کرده و نقاط حساس رشته‌ای و روندهای تکاملی متمایزی را آشکار ساخته است. در حوزه آموزشی، بائک و دولک (Baek & Doleck, 2022) و ناگندرا و جن (Nagendhra & Jen, 2024) افزایش شدید انتشارات حوزه داده‌کاوی را تأیید می‌کنند که تمرکز غالب آنها بر پیش‌بینی عملکرد تحصیلی دانش‌آموزان و تشخیص رفتارهای یادگیری با استفاده از تکنیک‌های پیشرفته است، اگرچه آنها به پراکندگی در ادبیات نیز اشاره می‌کنند. به طور مشابه، در حوزه مراقبت‌های بهداشتی و سلامت عمومی، مطالعات دوس سانتوس و همکاران (dos Santos et al., 2019)، هو و همکاران (Hu et al., 2020) و کولینگ و همکاران (Kolling et al., 2021) نشان می‌دهد که تحقیقات از سال ۲۰۱۷ به طور قابل توجهی تشدید شده و موضوعات حرکتی مانند شبکه‌های عصبی، سرطان، پرونده‌های الکترونیکی سلامت و دیابت شیرین را شناسایی کرده‌اند. این بررسی‌ها به طور مداوم ماشین‌های بردار پشتیبان (SVM) و زبان R را به عنوان ابزارهای رایج برجسته می‌کنند، در حالی که به مرزهای نوظهوری مانند پزشکی دقیق و التهاب نیز اشاره دارند. فراتر از آموزش و سلامت، عالم و همکاران (Alam et al., 2023) و پینادات و همکاران (Pynadath et al., 2022) به ترتیب تأثیر داده‌کاوی را در کتابخانه‌های

دیجیتال و مدیریت ارتباط با مشتری نشان می‌دهند، که مورد دوم به یک تغییر تکاملی آشکار به سمت چارچوب‌های یادگیری ماشین و مدیریت ارتباط با مشتری مبتنی بر هوش مصنوعی اشاره می‌کند.

در تکمیل این چشم‌انداز، سایر ارزیابی‌های کتاب‌سنجی، شکاف‌های بحرانی و الگوهای منحصر به فرد مختص به هر حوزه را شناسایی کرده‌اند. به عنوان مثال، دا سیلوا و همکاران (da Silva et al., 2020) شکاف قابل توجهی در ادبیات مربوط به کاربرد داده‌کاوی در مدیریت ریسک زنجیره تأمین یافتند و از طراحی مجدد فرآیندهای زنجیره تأمین برای ادغام این تکنیک‌ها حمایت کردند. در همین حال، اوپادیای (Upadhyay, 2018) یک تحلیل اولیه و بنیادی ارائه می‌دهد که همگرایی بین ادبیات داده‌کاوی و کلان‌داده را برجسته می‌کند و خاطرنشان می‌کند که بخش قابل توجهی از نویسندگان کلان‌داده از پیشینه داده‌کاوی سرچشمه می‌گیرند، با این حال تنها همپوشانی جزئی در کشورهای برتر و حوزه‌های تحقیقاتی وجود دارد.

در داخل کشور نیز در نزدیک‌ترین مطالعه به پژوهش حاضر، رفیعی خشنود و کرامت‌فر (۱۴۰۱) به مقایسه تحلیلی ۱۰۸۲ مدرک علمی ایران در حوزه موضوعی متن‌کاوی پرداختند و نتیجه گرفتند که میانگین نرخ رشد مدارک علمی ایران در حوزه متن‌کاوی بالاتر از حوزه‌های موضوعی دیگر است و کشورهای آمریکا، انگلیس و استرالیا بیشترین میزان مشارکت را با محققان ایرانی در این حوزه موضوعی داشته‌اند. همچنین مشخص شد مدارک علمی بین‌المللی که بر زبان انگلیسی متمرکز هستند، استناد بیشتری نسبت به مدارک علمی متمرکز بر زبان فارسی دریافت می‌کنند. حاجیان و زرچینی (۱۴۰۳) نیز در پژوهشی با رویکرد علم‌سنجی به تحلیل کاربردهای داده‌کاوی در صنعت بیمه پرداختند و نشان دادند که رشد داده‌ها و ضرورت شناسایی ریسک، شرکت‌های بیمه را به سمت روش‌های نوین داده‌کاوی سوق داده است. همچنین، به رتبه‌بندی کشورهای برتر در این حوزه و جایگاه ایران اشاره کرده‌اند.

از نظر روش‌شناسی، پیشینه‌های فوق به ابزارهایی مانند ووس–ویور ، بیبلیومتریکس ، و سایت اسپیس و متکی هستند و از تحلیل‌های هم‌رخدادی، تحلیل هم‌استنادی و نمودارهای راهبردی برای ترسیم ساختارهای موضوعی استفاده کرده‌اند. در مجموع، این ادبیات تأکید می‌کند که تحقیقات همچنان در علوم کامپیوتر و حوزه‌های کاربردی خاص متمرکز است و فرصت‌های مداومی برای همکاری بین رشته‌ای در حوزه‌های کمتر بررسی‌شده و در پیوند دادن تکنیک‌های داده‌کاوی با نظریه‌های خاص حوزه، وجود دارد.

بررسی دقیق پیشینه، خلأهای پژوهشی را نمایان می‌سازد. به ویژه آنکه، از نظر داده‌ای، عمده مطالعات علم‌سنجی موجود در حوزه داده‌کاوی و علوم داده، از پایگاه‌های بین‌المللی مانند وب‌آوساینس و اسکوپوس استفاده کرده‌اند و هیچ مطالعه‌ای به‌طور اختصاصی تولیدات علمی فارسی‌زبان این حوزه را تحلیل نکرده است. این غفلت، منجر به نادیده ماندن پویایی‌های بومی این حوزه شده است. به طوری که، هیچ مطالعه‌ای روند تکاملی پژوهش‌های فارسی داده‌کاوی را در یک افق بلندمدت ۲۰ ساله (۱۳۸۴–۱۴۰۳) بررسی نکرده است و بنابراین چرخه حیات مفاهیم در این حوزه مشخص نبوده است. پژوهش حاضر با پوشش هم‌زمان این شکاف (داده‌ای و زمانی) و با بهره‌گیری از روش‌های علم‌سنجی، درصدد است تا تصویری جامع از ساختار مفهومی پژوهش‌های فارسی حوزه داده‌کاوی ارائه دهد.

روش‌شناسی پژوهش

پژوهش حاضر از نوع توصیفی با رویکرد کاربردی و مبتنی بر تحلیل علم سنجی است که با بهره‌گیری از تحلیل هم‌رخدادی واژگان، به بررسی ساختار مفهومی و تکامل موضوعی پژوهش‌های فارسی حوزه داده‌کاوی پرداخته است. جامعه پژوهش شامل تمامی مقالات فارسی نمایه‌شده در نمایه استنادی علوم ایران وابسته به مؤسسه استنادی و پایش علم و فناوری جهان اسلام (آی‌اس‌سی) بود. فرایند بازیابی اطلاعات در فروردین‌ماه ۱۴۰۵ انجام شد. راهبرد جستجو بر مبنای جستجو در فیلد عنوان مقالات و با استفاده از دو صورت نگارشی رایج اصطلاح «داده‌کاوی» و «داده کاوی» در میان تمامی نشریات نمایه‌شده فارسی پایگاه آی‌اس‌سی طراحی شد. بازه زمانی جستجو کلیه مقالات منتشرشده از سال ۱۳۸۴ تا ۱۴۰۳ را در بر گرفت.

نتیجه جستجو منجر به بازیابی ۵۹۵ رکورد شد. پس از بررسی دستی عنوان، چکیده و در موارد لازم متن مقاله توسط متخصص موضوعی (یکی از نویسندگان که دارای مدرک دکتری در رشته مهندسی کامپیوتر است)، تعداد ۷ مقاله به دلیل عدم ارتباط موضوعی

با حوزه داده‌کاوی از مجموعه حذف شدند و در نهایت ۵۸۹ مقاله به‌عنوان مجموعه نهایی تحلیل انتخاب شدند.

از آنجا که پایگاه آی‌اس‌سی امکان استخراج اطلاعات در قالب فایل‌های استاندارد موردنیاز نرم‌افزارهای علم‌سنجی را فراهم نمی‌کند، اطلاعات کتاب‌شناختی هر مقاله شامل عنوان، نویسندگان، سال انتشار، نام نشریه و کلیدواژه‌های نویسندگان به‌صورت دستی استخراج و ابتدا در محیط اکسل سازمان‌دهی شد. سپس، به‌منظور جلوگیری از ایجاد خطاهای ناشی از تفاوت‌های نگارشی در پردازش خودکار داده‌ها، فرایند یکسان‌سازی نویسه‌های فارسی و عربی انجام گرفت؛ به‌گونه‌ای که نویسه‌های هم‌ارز مانند «ی/ی»، «ک/ک» و همچنین تفاوت‌های ناشی از استفاده از نیم‌فاصله و فاصله در واژگان استانداردسازی شدند. در ادامه، داده‌ها به‌صورت دستی در قالب ساختاری سازگار با قالب استاندارد وب‌آوساینس سازمان‌دهی شدند تا قابلیت پردازش توسط نرم‌افزار بیبلیوشانی را داشته باشد.

پیش از انجام تحلیل، کلیدواژه‌ها به‌صورت دستی توسط متخصص موضوعی مورد بازبینی و استانداردسازی قرار گرفتند تا از ایجاد پراکندگی مصنوعی در شبکه هم‌رخدادی جلوگیری شود. در این مرحله، شکل‌های مختلف نگارشی واژگان یکسان‌سازی، خطاهای تایپی اصلاح، اختلاف‌های ناشی از فاصله و نیم‌فاصله برطرف، صورت‌های مفرد و جمع ادغام و کلیدواژه‌های مترادف یا دارای مفهوم یکسان استاندارد شدند. برای نمونه، کلیدواژه‌های «شبکه عصبی» و «شبکه‌های عصبی»، «طبقه‌بندی» و «رده‌بندی»، «قانون انجمنی» و «قواعد انجمنی»، «پایگاه داده» و «پایگاه‌های داده» و همچنین «درخت تصمیم» و «درخت تصمیم‌گیری» در قالب یک اصطلاح استاندارد ادغام شدند. ذکر این نکته ضروری است که با توجه به اینکه تمام مدارک بازیابی شده به صراحت به داده‌کاوی مرتبط بودند، کلیدواژهٔ «داده‌کاوی»، به عنوان کلیدواژهٔ ناظر بر عام، از تحلیل حذف شد. این تصمیم با پیروی از شیوه‌های کتاب‌سنجی مرسوم است که اصطلاحات بسیار کلی که در اکثر اسناد ظاهر می‌شوند و به تمایز موضوعی کمکی نمی‌کنند، حذف می‌شوند تا شناسایی مضامین تحقیقاتی و ساختارهای مفهومی زیربنایی بهبود یابد.

تحلیل داده‌ها با استفاده از بسته بیبلیومتریک و رابط وب «بیبلیوشاینی » در محیط نرم‌افزار آر انجام شد. بیبلیوشاینی از ابزارهای کاربرپسند و رایجی است که استفاده از آن در مطالعات علم‌سنجی رو به رواج است (Aria & Cuccurullo, 2017; Aria et al., 2026). در مرحلهٔ بعد با توجه به امکاناتی که در نرم‌افزار بیبلیوشاینی وجود دارد نقشه‌های علمی این حوزه ترسیم به پرسش‌های پژوهش پاسخ داده شد.

به منظور انجام تحلیل هم واژگانی، پس از بررسی فراوانی کلیدواژه‌ها، تعداد ۶۳ کلیدواژه که حداقل ۳ بار در مجموعه داده‌ها تکرار شده بودند وارد تحلیل شدند تا میزان هم‌رخدادی آنان در مقاله‌ها به‌دست آید و سپس شبکه هم واژگانی و نمودار راهبردی حوزهٔ داده کاوی با استفاده از بیبلیوشاینی ترسیم شد. با توجه به کم بودن تعداد کلیدواژه ها، پژوهشگران ناچار بودند حد آستانه را به فراوانی ۳ تقلیل دهند. در قسمت تحلیل تکامل موضوعی نیز بازه زمانی پژوهش به سه دوره ۱۳۸۴–۱۳۹۵، ۱۳۹۶–۱۳۹۹ و ۱۴۰۰–۱۴۰۳ تقسیم شد. این تقسیم‌بندی با هدف ایجاد تعادل نسبی در حجم مدارک هر دوره و فراهم کردن امکان مقایسه معنادار مضامین در طول زمان انجام شد (Ho & Lwesya, 2025)؛ زیرا در سال‌های آغازین، تعداد مقالات اندک بود و استفاده از بازه‌های کوتاه‌تر به تشکیل شبکه‌های مفهومی پایدار منجر نمی‌شد.

همچنین لازم به ذکر است که علیرغم سعی در حفظ جامعیت رویکرد پژوهش حاضر، چندین محدودیت اساسی باید در تفسیر یافته‌ها مدنظر قرار گیرند؛ محدودیت پایگاه داده یکی از آن هاست. تمرکز انحصاری بر پایگاه آی‌اس‌سی، هرچند مزیت بررسی ادبیات معتبرِ فارسی‌زبان را فراهم می‌کند، اما بخشی از تولیدات علمی پژوهشگران ایرانی که در مجلات بین‌المللی (اسکوپوس یا وب‌آوساینس) منتشر شده‌اند را از دایره بررسی خارج می‌سازد. بنابراین یافته‌ها بازتاب‌دهنده وضعیت پژوهش‌های فارسی‌زبان داخلی هستند و نه تمام تولیدات علمی پژوهشگران ایرانی در حوزه داده‌کاوی. همچنین، داده‌های مربوط به سال ۱۴۰۴ هنوز در پایگاه آی‌اس‌سی نمایه‌سازی نشده است و این امر می‌تواند روندهای موضوعی اخیر را کمتر از واقعیت نشان دهد. از دیگر محدودیت های این مطالعه که به پایگاه‌داده باز می‌گردد اطلاعات مربوط به تحلیل استنادی است که به دلیل نقص در اطلاعات استنادی موجود در پایگاه

آی‌اس‌سی، تحلیل هم‌استنادی و استنادات محوری که اطلاعات ارزشمندی درباره تأثیرگذاری علمی فراهم می‌کنند، در این پژوهش امکان‌پذیر نبود. این محدودیت، ارائهٔ تصویر کامل‌تری از تأثیر علمی مقالات را ناممکن می‌سازد.

یافته‌های پژوهش

پاسخ به پرسش اول پژوهش: تکامل موضوعی و مسیر تغییرات موضوعات پژوهشی در حوزه‌ی داده‌کاوی در گذر زمان چگونه بوده است؟

تجزیه و تحلیل داده‌ها نشان داد که روی هم رفته تعداد ۱۷۳۴ کلیدواژه در ۵۸۹ مقالهٔ تحت مطالعه استفاده شده است که معادل میانگین ۳ کلیدواژه به ازای هر مقاله است. فراوانی تجمعی ۱۰ کلیدواژه برتر از سال ۱۳۸۶ تا ۱۴۰۳ که از نظر تحلیل تکامل موضوعی در گذر زمان، اهمیت وافری دارد در شکل ۱ نشان داده شده است.

بررسی دقیق‌تر این شکل حاکی از آن است که «درخت تصمیم» با ۷۸ رخداد، بیشترین فراوانی تجمعی را در کل دوره دارد و از نظر تکرار کلیدواژه، در جایگاه نخست قرار گرفته است. افزایش فراوانی تجمعی این کلیدواژه از حدود سال ۱۳۹۲ محسوس‌تر شده و تا پایان دوره ادامه یافته است. با توجه به ماهیت داده، این یافته را می‌توان نشان‌دهنده حضور و تکرار بیشتر مفهوم «درخت تصمیم» در مجموعه مورد بررسی دانست؛ با این حال، فراوانی یک کلیدواژه به‌تنهایی بیانگر میزان استفاده واقعی از روش، اهمیت علمی یا برتری آن نسبت به سایر روش‌ها نیست.

همچنین «شبکه عصبی مصنوعی» با ۷۴ رخداد، «خوشه‌بندی» با ۵۳ رخداد، «پیش‌بینی» با ۴۳ رخداد و «ماشین بردار پشتیبان» با ۳۵ رخداد از دیگر کلیدواژه‌های پرتکرار مجموعه هستند. این الگو نشان می‌دهد که واژگان مورد استفاده در ادبیات داده‌کاوی مورد بررسی، در بخش قابل توجهی حول مجموعه‌ای از مفاهیم و روش‌های روش‌شناختی مشخص شکل گرفته است و این مفاهیم در بخش‌های مختلف دوره مورد مطالعه تداوم داشته‌اند.

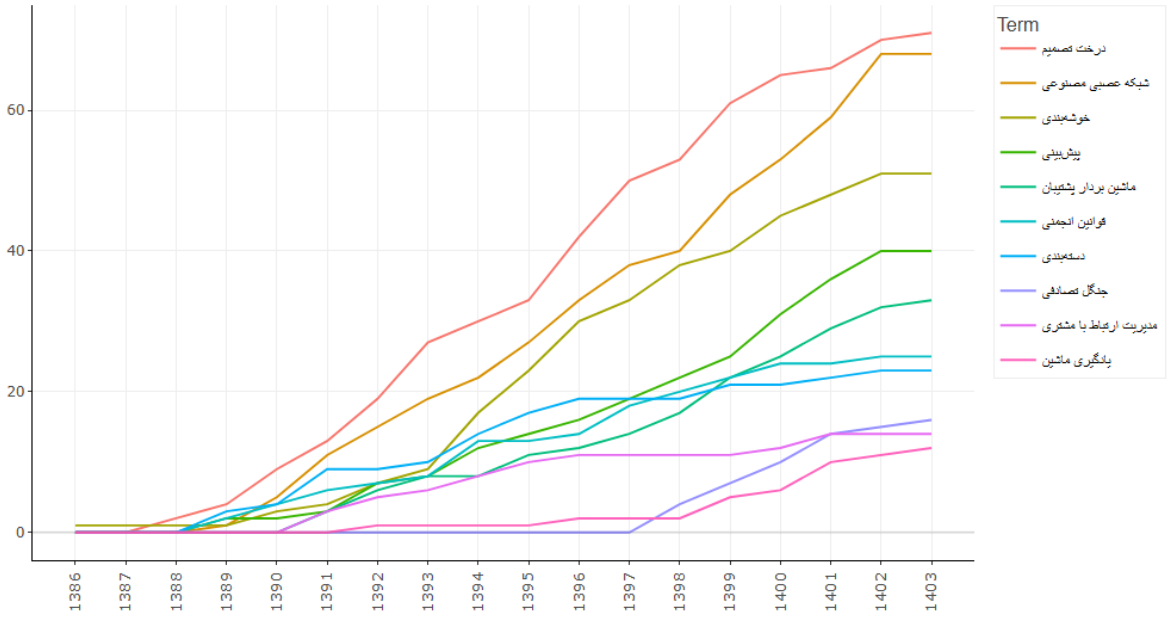
از منظر زمانی، نکته مهم صرفاً افزایش یا کاهش یک کلیدواژه منفرد نیست، بلکه هم‌زمانی تداوم برخی مفاهیم با افزایش حضور برخی اصطلاحات در سال‌های پایانی است. برای مثال، «درخت تصمیم»، «شبکه عصبی مصنوعی» و «خوشه‌بندی» در بخش‌های مختلف دوره حضور دارند، در حالی که «جنگل تصادفی» و «یادگیری ماشین» در سال‌های پایانی بیشتر مشاهده می‌شوند. بنابراین، الگوی زمانی مجموعه بیشتر از آنکه بیانگر حذف یک دسته از روش‌ها و جایگزینی کامل آن‌ها با دسته‌ای دیگر باشد، تداوم برخی مفاهیم روش‌شناختی را در کنار تغییر تدریجی در ترکیب واژگان مورد استفاده در دوره‌های مختلف نشان می‌دهد.

الگوی مشاهده‌شده در فراوانی کلیدواژه‌ها، در روند تحول موضوعات ارائه‌شده در شکل ۲ نیز قابل پیگیری است. مقایسه دو شکل نشان می‌دهد که تغییرات موضوعی در دوره مورد بررسی عمدتاً با تداوم برخی موضوعات و هم‌زمان ظهور یا افزایش حضور برخی موضوعات دیگر همراه بوده است.

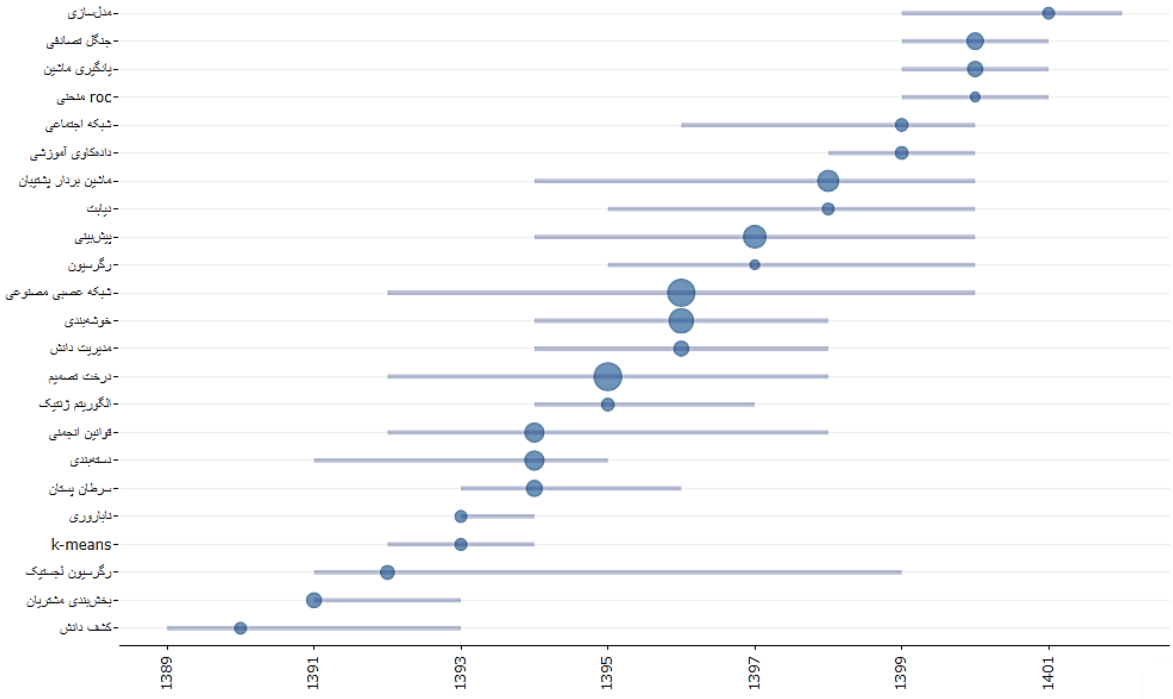
از این منظر، فراوانی بالای «درخت تصمیم» نشان‌دهنده حضور و تکرار بیشتر این کلیدواژه در مجموعه مورد مطالعه است، در حالی که افزایش حضور «یادگیری ماشین» و «جنگل تصادفی» در دوره‌های پایانی، تغییر در واژگان موضوعی و روش‌شناختی مورد استفاده در ادبیات بازیابی شده را نشان می‌دهد. به‌ویژه، «یادگیری ماشین» مفهومی در سطحی کلی‌تر از یک الگوریتم مشخص است؛ بنابراین افزایش حضور آن را نمی‌توان به‌تنهایی به معنای رواج یک الگوریتم خاص یا حرکت قطعی پژوهش‌ها به سمت روش‌های پیشرفته‌تر تفسیر کرد.

در مجموع، الگوی زمانی کلیدواژه‌ها دو ویژگی مکمل را نشان می‌دهد: نخست، تداوم حضور مجموعه‌ای از مفاهیم روش‌شناختی پرتکرار در طول دوره؛ و دوم، افزایش یا ورود برخی اصطلاحات در سال‌های پایانی. بنابراین، روند مشاهده شده را می‌توان نه به صورت جایگزینی خطی روش‌های قدیمی با روش‌های جدید، بلکه به‌صورت تداوم یک هسته روش‌شناختی همراه با تغییر و بازترکیب تدریجی

واژگان موضـوعی در طول زمان تفسـیر کرد. بررسـی دقیق‌تر روابط میان این مفاهیم و نحوه تداوم، انشـعاب و ارتباط موضـوعات در دوره‌های مختلف، در تحلیل تکامل موضوعی پژوهش‌های حوزه داده‌کاوی در ادامه بررسی می‌شود.

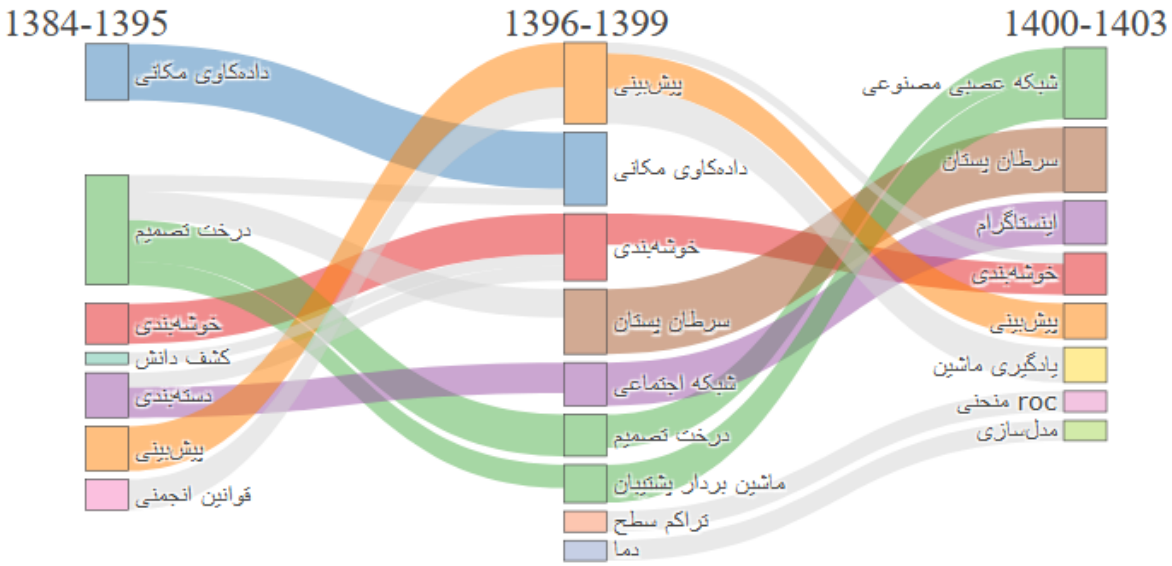


شکل ۱: کلیدواژه های برتر در مقالات داده کاوی در گذر زمان



شکل ۲: روند تحول موضوعات مورد توجه در مقالات حوزهٔ داده کاوی

تحلیل تکامل مو ضوعی در حوزهٔ داده‌کاوی که ن شان‌دهنده‌ی م سیر تغییرات و تداوم مو ضوعات پژوه شی در سه بازهٔ زمانی مختلف (۱۳۸۴–۱۳۹۵، ۱۳۹۶–۱۳۹۹ و ۱۴۰۰–۱۴۰۳) است در شکل ۳ نشان داده شده است. این تقسیم‌بندی زمانی با هدف ایجاد تعادل نسبی در تعداد مدارک هر دوره و افزایش قابلیت مقایسـه سـاختارهای مفهومی انجام شـد. این تحلیل در واقع مسـیر جریان دانش و میزان پایداری مضامین در حوزهٔ داده‌کاوی را در گذار از دوره‌های زمانی مختلف (۱۳۸۴ تاکنون) نشان می‌دهد.



شکل ۳: نقشهٔ تکامل موضوعی حوزهٔ داده‌کاوی در گذر زمان

تحلیل تکامل موضوعی در سه بازه زمانی (۱۳۸۴–۱۳۹۵، ۱۳۹۶–۱۳۹۹ و ۱۴۰۰–۱۴۰۳)، علاوه بر تداوم برخی مضامین، تغییر در ترکیب و نحوه همراهی کلیدواژه‌ها را نشان می‌دهد. «درخت تصمیم» و «خوشه‌بندی» در بیش از یک دوره حضور دارند و در نتیجه می‌توان آن‌ها را از جمله مضامینی دانست که در طول دوره مطالعه پیوستگی بیشتری در مجموعه مورد بررسی داشته‌اند. در مقابل، جریان‌های مشاهده‌شده از «درخت تصمیم» در دوره دوم به «ماشین بردار پشتیبان» و در دوره سوم به «شبکه عصبی مصنوعی» نشان می‌دهد که این کلیدواژه‌ها در دوره‌های متوالی در ساختار موضوعی مجموعه با یکدیگر مرتبط بوده‌اند.

نکته مهم در تفسیر این جریان‌ها آن است که اتصال یک کلیدواژه از یک دوره به کلیدواژه‌ای در دوره بعد الزاماً به معنای جایگزینی روش اول با روش دوم نیست؛ بلکه نشان می‌دهد این مفاهیم در ساختار همرخدادی دوره‌های مختلف، تداوم یا ارتباط موضوعی داشته‌اند. بنابراین، الگوی مشاهده‌شده را بهتر است به‌صورت تداوم برخی روش‌های شناخته‌شده همراه با بازترکیب و انشعاب موضوعات در دوره‌های بعدی تفسیر کرد.

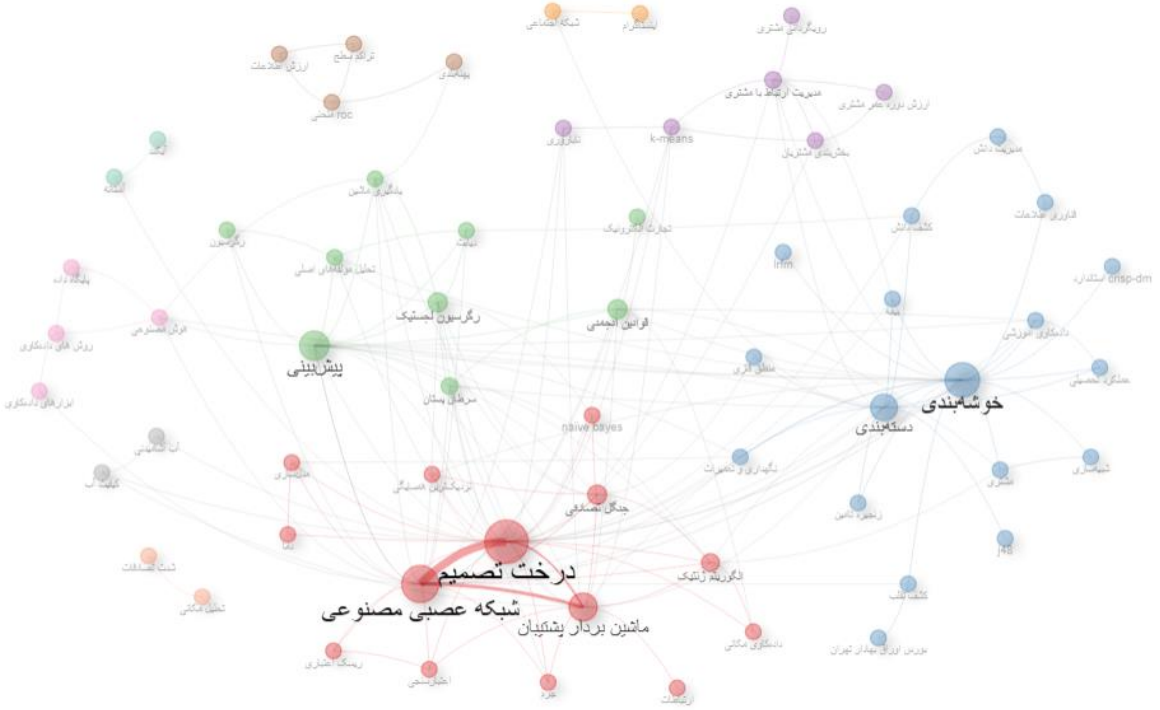
از سوی دیگر، ارتباط «خوشه‌بندی» با «سرطان پستان» نمونه‌ای از پیوند میان یک روش داده‌کاوی و یک حوزه کاربردی مشخص است. به بیان دیگر، موضوعاتی مانند «سرطان پستان»، «شبکه اجتماعی» و «دما» در این تحلیل به‌عنوان حوزه‌های کاربردی به‌کارگیری روش‌های داده‌کاوی ظاهر شده‌اند. حضور این نوع جریان‌ها نشان می‌دهد که ساختار موضوعی پژوهش‌های مورد بررسی فقط حول روش‌ها شکل نگرفته، بلکه در برخی دوره‌ها روش‌های داده‌کاوی در ارتباط با مسائل کاربردی مشخص نیز ظاهر شده‌اند.

در مجموع، شکل ۳ بیش از آنکه یک مسیر خطی از «روش‌های قدیمی» به «روش‌های جدید» را نشان دهد، سه ویژگی را همزمان آشکار می‌کند: تداوم برخی مضامین روش‌شناختی، بازترکیب و انشعاب میان مضامین در دوره‌های متوالی، و پیوند برخی روش‌ها با زمینه‌های کاربردی مشخص.

پاسخ به پرسش دوم پژوهش: ساختار مفهومی در حوزهٔ داده‌کاوی، بر اساس تحلیل شبکهٔ هم‌رخدادی کلیدواژه‌ها چگونه است و خوشه‌های موضوعی اصلی در این شبکه کدامند؟

به منظور درک عمیق‌تر از ساختار دانش و پیوندهای مفهومی در حوزهٔ داده‌کاوی، شبکهٔ هم‌رخدادی کلیدواژه‌ها ترسیم شد. همانطور که در بخش روش شناسی ذکر شد، تعداد ۶۳ کلیدواژه (گره) با حداقل فراوانی ۳ که مشتمل بر ۳۲۷ یال بودند وارد تحلیل هم واژگانی شدند. این فرایند با استفاده از امکانات نهادینه شده در نرم افزار بیبلیوشاینی صورت گرفت که در نهایت منجر به ترسیم شبکهٔ

هم‌رخدادی کلیدواژه‌ها در حوزهٔ داده‌کاوی شد. این شبکه که در شکل ۴ به تصویر کشیده شده است، خوشه‌های پژوهشی شکل گرفته در حوزهٔ داده‌کاوی را نشان می‌دهد. این ساختار شبکه‌ای، زیربنای اصلی نقشهٔ موضوعی را تشکیل می‌دهد که در بخش بعدی بدان پرداخته شده است. در واقع، خوشه‌های شناسایی شده در این شبکه، همان واحدهای تحلیل در نقشهٔ موضوعی هستند که بر اساس قدرت پیوندهای داخلی (چگالی) و تعاملات میان‌گروهی (مرکزیت)، در قسمتهای چهارگانهٔ نمودار راهبردی قرار گرفته‌اند.



شکل ۴: شبکهٔ هم‌رخدادی کلیدواژه‌ها در حوزهٔ داده‌کاوی

تحلیل هم‌رخدادی کلیدواژه‌ها در حوزهٔ داده‌کاوی مشتمل بر ۱۰ خوشه مفهومی است. با توجه به شکل ۴، می‌توان یک هسته روش‌شناختی نسبتاً پررنگ و در کنار آن، مجموعه‌ای از خوشه‌های تخصصی‌تر را مشاهده کرد. «درخت تصمیم»، «شبکه عصبی مصنوعی»، «خوشه‌بندی» و «پیش‌بینی» از مهم‌ترین گره‌های این ساختار هستند و بخش قابل توجهی از ارتباطات شبکه پیرامون این دسته از مفاهیم شکل گرفته است.

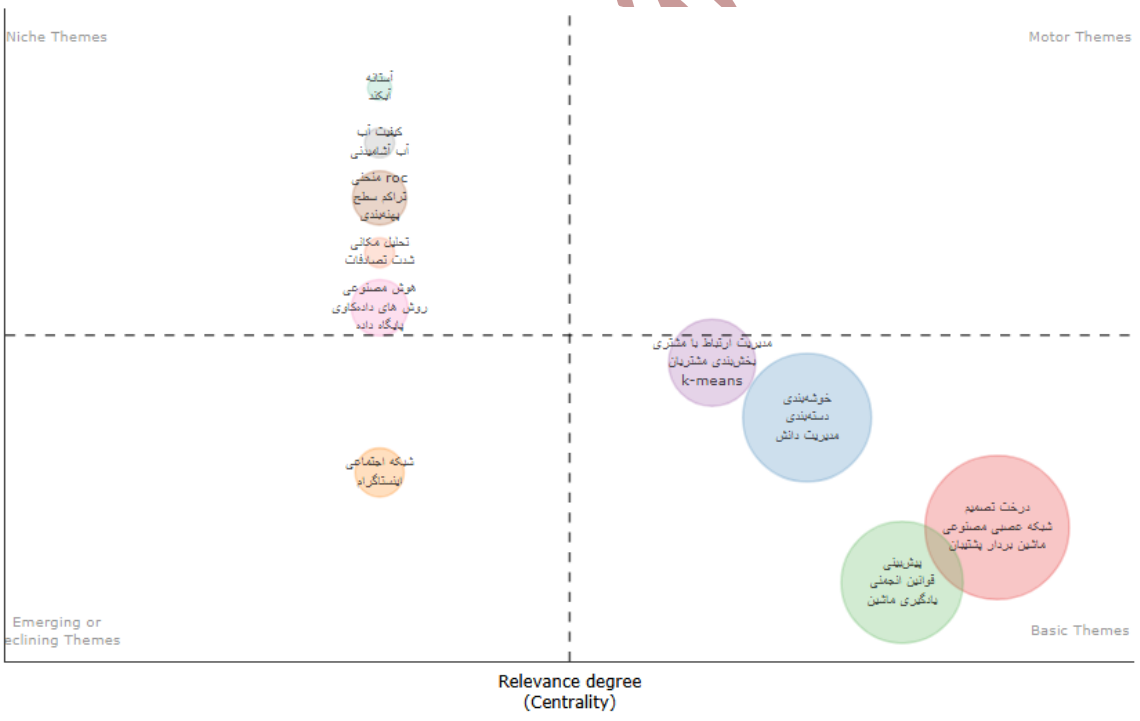
این الگو نشان می‌دهد که ساختار مفهومی مجموعه مورد بررسی صرفاً از موضوعات کاربردی منفرد تشکیل نشده، بلکه دارای یک لایه روش‌شناختی مرکزی است که با مجموعه‌ای از زمینه‌های کاربردی و تخصصی ارتباط پیدا می‌کند. در کنار این هسته، خوشه‌هایی شامل موضوعاتی مانند «کیفیت آب»، «تحلیل مکانی» و «مدیریت ارتباط با مشتری» قرار گرفته‌اند که نشان می‌دهند روش‌های داده‌کاوی در ادبیات مورد بررسی در مسائل مشخص و متفاوتی به کار رفته‌اند.

بنابراین، شبکه را می‌توان به‌صورت یک ساختار چندلایه تفسیر کرد: در لایه نخست، روش‌ها و مفاهیم عمومی‌تر داده‌کاوی قرار دارند؛ در لایه دوم، خوشه‌های تخصصی و کاربردی پیرامون این مفاهیم شکل گرفته‌اند. اهمیت این یافته در آن است که نشان می‌دهد پیوند مفهومی میان پژوهش‌های داده‌کاوی در مجموعه فارسی مورد بررسی، هم از مسیر اشتراک روش‌ها و هم از مسیر کاربرد روش‌ها در مسائل مختلف شکل گرفته است.

پاسخ به پرسش سوم پژوهش: جایگاه و ساختار موضوعات پژوهش‌های فارسی حوزه داده‌کاوی بر اساس نمودار راهبردی چگونه است؟

نقشه موضوعی، توصیف ارتباط درونی و همبستگی بین خوشه‌های موضوعی را نشان می‌دهد. در این نمودار، اغلب از محور افقی برای ارائه مرکزیت (میزان همبستگی خوشه‌ها) و از محور عمودی برای ارائه تراکم (میزان توان ارتباط درونی هر خوشه) استفاده می‌شود (سهیلی و همکاران، ۱۳۹۷). این نقشه، که از آن به نمودار راهبردی یاد می‌شود، موضوعات تحقیقاتی را به چهار ربع طبقه‌بندی می‌کند: ۱. موضوعات محرک (بالا سمت راست): که مرکزیت بالا و تراکم بالا دارند. این موضوعات به خوبی توسعه یافته و برای حوزه بسیار مهم هستند و آنها «موتورهای» حوزه تحقیقاتی هستند؛ ۲. موضوعات تخصصی (بالا سمت چپ): که مرکزیت پایین و تراکم بالا دارند. این خوشه‌ها بسیار توسعه یافته اما منزوی و حاشیه‌ای هستند (تحقیقات تخصصی)؛ ۳. موضوعات نوظهور یا رو به زوال (پایین سمت چپ): که مرکزیت پایین و تراکم پایین دارند. این موضوعات یا تازه شروع به ظهور کرده‌اند یا در حال محو شدن هستند؛ و ۴. مضامین پایه/عرضی (پایین سمت راست): که مرکزیت بالا و تراکم پایینی دارند. این مضامین برای این حوزه مهم و مرکزی هستند، ولی ممکن است به خوبی در درون توسعه نیافته باشند (آنها به عنوان مفاهیم بنیادی عمل می‌کنند).

نمودار راهبردی مربوط به نقشه موضوعی پژوهش‌های حوزه داده‌کاوی در شکل ۵ نشان داده شده است. بررسی این نمودار نشان می‌دهد که خوشه‌های موضوعی در ربع‌های مختلف پراکنده شده‌اند، اما توزیع آن‌ها یکسان نیست. از این رو، برای تفسیر ساختار موضوعی مجموعه مورد بررسی، جایگاه خوشه‌ها در هر یک از چهار ربع و معنای مرکزیت و تراکم آن‌ها مورد توجه قرار گرفته است.



شکل ۵: نمودار راهبردی مربوط به نقشهٔ موضوعی پژوهش های حوزهٔ داده کاوی

نخست، نکته قابل توجه عدم مشاهده خوشه‌ای در ربع بالا سمت راست، یعنی بخش مضامین محرک است. این وضعیت نشان می‌دهد که در شبکه مورد بررسی، خوشه‌ای که به طور هم‌زمان از مرکزیت و تراکم بالایی برخوردار باشد مشاهده نشده است. با این حال، این یافته را نمی‌توان به تنهایی به معنای عدم بلوغ یا ضعف حوزه داده‌کاوی در ایران دانست؛ زیرا نتیجه نمودار به ویژگی‌های شبکه مورد استفاده، از جمله حداقل فراوانی سه رخداد، انتخاب ۶۳ کلیدواژه و نحوه پراکندگی و هم‌رخدادی کلیدواژه‌ها وابسته است. بنابراین، نبود

خوشه در این ربع بیشتر بیانگر ساختار شبکه موضوعی در جامعه مورد مطالعه و تحت شرایط تحلیل حاضر است.

در مقابل، بخشی از موضوعات در ربع پایین سمت راست، یعنی مضامین پایه، قرار گرفته‌اند. این بخش شامل موضوعاتی مانند «درخت تصمیم»، «شبکه عصبی مصنوعی» و «ماشین بردار پشتیبان» است. قرار گرفتن این موضوعات در ناحیه با مرکزیت بالاتر نشان می‌دهد که آن‌ها با طیف متنوع‌تری از سایر موضوعات شبکه ارتباط دارند؛ در حالی که تراکم پایین‌تر آن‌ها نشان می‌دهد که پیوندهای درونی این خوشه‌ها نسبت به خوشه‌های واقع در نواحی پرتراکم محدودتر است. بنابراین، این موضوعات در ساختار شبکه حاضر، نقش ارتباط‌دهنده و پایه‌ای دارند و در ارتباط میان بخش‌های مختلف موضوعی شبکه حضور گسترده‌تری نشان می‌دهند. البته این جایگاه به معنای برتری علمی یا کارآمدی ذاتی این روش‌ها نیست، بلکه بیانگر موقعیت آن‌ها در ساختار هم‌رخدادی مجموعه مورد بررسی است.

در ربع بالا سمت چپ، یعنی موضوعات تخصصی، مفاهیمی مانند «کیفیت آب» و «تحلیل مکانی» قرار گرفته‌اند. تراکم بالاتر این موضوعات بیانگر انسجام بیشتر روابط درونی آن‌هاست، در حالی که مرکزیت پایین‌تر نشان می‌دهد ارتباط آن‌ها با سایر خوشه‌های شبکه محدودتر است. از این رو، این موضوعات را می‌توان در مجموعه مورد بررسی، موضوعاتی تخصصی و نسبتاً متمرکز دانست که ارتباطات درونی آن‌ها نسبتاً منسجم‌تر است، اما پیوند کمتری با سایر بخش‌های ساختار موضوعی شبکه دارند.

در ربع پایین سمت چپ نیز موضوعاتی مانند «شبکه اجتماعی» و «اینستاگرام» قرار گرفته‌اند. این موضوعات از مرکزیت و تراکم پایین‌تری برخوردارند؛ بنابراین، در شبکه مورد بررسی هم ارتباط آن‌ها با سایر خوشه‌ها محدودتر است و هم انسجام درونی کمتری دارند. با این حال، قرار گرفتن در این ربع به تنهایی نمی‌تواند مشخص کند که یک موضوع واقعاً در حال ظهور یا زوال است. به بیان دیگر، برچسب «نوظهور یا رو به زوال» در نمودار راهبردی هر دو وضعیت را دربرمی‌گیرد و برای تعیین جهت زمانی تغییرات به شواهد حاصل از تحلیل‌های زمانی نیاز است. بنابراین، درباره نوظهور یا رو به زوال بودن «شبکه اجتماعی» و «اینستاگرام» نمی‌توان صرفاً بر اساس موقعیت آن‌ها در شکل ۵ داوری قطعی کرد.

در مجموع، نمودار راهبردی نشان می‌دهد که ساختار موضوعی پژوهش‌های فارسی داده‌کاوی در مجموعه مورد بررسی، بیش از همه با حضور مضامین پایه و موضوعات تخصصی مشخص می‌شود. موضوعاتی مانند «درخت تصمیم»، «شبکه عصبی مصنوعی» و «ماشین بردار پشتیبان» به دلیل مرکزیت بالاتر، با بخش‌های مختلف شبکه ارتباط بیشتری دارند و در ساختار موضوعی نقش پایه‌ای و ارتباط‌دهنده ایفا می‌کنند؛ در مقابل، موضوعاتی مانند «کیفیت آب» و «تحلیل مکانی» با تراکم بالاتر و مرکزیت پایین‌تر، ساختاری تخصصی‌تر و متمرکزتر نشان می‌دهند. همچنین، قرار گرفتن «شبکه اجتماعی» و «اینستاگرام» در ربع پایین سمت چپ، جایگاه کم‌مرکز و کم‌تراکم آن‌ها را در شبکه حاضر نشان می‌دهد، اما جهت زمانی تغییرات آن‌ها را مشخص نمی‌کند.

بحث و نتیجه‌گیری

پژوهش حاضر با هدف بررسی روند موضوعات، تکامل موضوعی، ساختار هم‌رخدادی و جایگاه موضوعات پژوهش‌های فارسی حوزه داده‌کاوی در مقالات نمایه‌شده در پایگاه آی‌اس‌سی طی سال‌های ۱۳۸۴ تا ۱۴۰۳ انجام شد. تحلیل ۵۸۹ مقاله و ۱۷۳۴ کلیدواژه، در مجموع الگویی را نشان داد که در آن تداوم برخی مفاهیم روش‌شناختی پرتکرار با تغییر تدریجی در ترکیب موضوعات و شکل‌گیری پیوندهای متفاوت میان موضوعات روش‌شناختی و کاربردی همراه بوده است. بنابراین، الگوی مشاهده‌شده را نمی‌توان صرفاً به افزایش یا کاهش فراوانی چند کلیدواژه تقلیل داد؛ بلکه بررسی هم‌زمان روند زمانی، روابط موضوعی و ساختار شبکه نشان می‌دهد که تغییرات موضوعی در این مجموعه بیشتر ماهیتی تدریجی و ترکیبی داشته‌اند.

از نظر روند زمانی، «درخت تصمیم» با ۷۸ رخداد و «شبکه عصبی مصنوعی» با ۷۴ رخداد پرتکرارترین کلیدواژه‌ها بودند و پس از آن‌ها «خوشه‌بندی»، «پیش‌بینی» و «ماشین بردار پشتیبان» قرار داشتند. تداوم حضور این مفاهیم در دوره‌های مختلف نشان می‌دهد که بخشی از واژگان روش‌شناختی در کل دوره مورد بررسی ثبات نسبی داشته‌اند. در کنار این تداوم، افزایش حضور اصطلاحاتی مانند

«جنگل تصادفی» و «یادگیری ماشین» در سال‌های پایانی بیانگر تغییر تدریجی در ترکیب واژگان موضوعی است. بنابراین، شواهد حاصل از روند زمانی بیشتر از یک الگوی تداوم همراه با تغییر و بازترکیب موضوعات حمایت می‌کند تا اینکه یک الگوی ساده از حذف روش‌های پیشین و جایگزینی کامل آن‌ها با روش‌های جدید باشد. از جانب دیگر، غیاب کلیدواژه‌هایی نظیر «یادگیری عمیق»، و «هوش مصنوعی قابل توضیح» در بین کلیدواژه های پرتکرار، در مقایسه با ادبیات بین‌المللی (Aimbetova et al., 2025)، حاکی از آن است که خلأهای روش‌شناختی مهمی در پژوهش‌های داخلی وجود دارد که نیاز به توجه جدی دارند.

تحلیل تکامل موضوعی این برداشت را تکمیل می‌کند. حضور «درخت تصمیم» و «خوشه‌بندی» در بیش از یک دوره و مشاهده جریان‌هایی میان موضوعات روش‌شناختی و کاربردی نشان می‌دهد که برخی مفاهیم در طول دوره حفظ شده‌اند، اما در دوره‌های مختلف با موضوعات دیگری ارتباط پیدا کرده‌اند. برای نمونه، ارتباط‌های مشاهده‌شده میان «درخت تصمیم» و روش‌هایی مانند «ماشین بردار پشتیبان» و «شبکه عصبی مصنوعی» و همچنین ارتباط «خوشه‌بندی» با «سرطان پستان»، نشان می‌دهد که تغییر موضوعی فقط از طریق ورود کلیدواژه‌های جدید رخ نداده، بلکه از طریق شکل‌گیری و تغییر ارتباط میان مفاهیم روش‌شناختی و حوزه‌های کاربردی نیز صورت گرفته است. همچنین، ظهور موضوعاتی مانند «شبکه اجتماعی» و «اینستاگرام» در سال های اخیر نشان می‌دهد که کاربرد روش‌های داده‌کاوی در پژوهش‌های فارسی مورد بررسی به حوزه‌های مرتبط با رسانه‌های اجتماعی نیز گسترش یافته است. این مشاهده با گزارش برخی مطالعات بین‌المللی مبنی بر گسترش کاربرد روش‌های داده‌کاوی در حوزه‌های مختلف، از جمله رسانه‌های اجتماعی، همسو است (Kousis & Tjortjis, 2021). بر این اساس، می‌توان گفت ساختار موضوعی پژوهش‌ها در دوره مورد بررسی از یک هسته نسبتاً پایدار از مفاهیم روش‌شناختی تشکیل شده که در کنار آن، پیوندها و کاربردهای موضوعی متفاوتی در دوره‌های مختلف شکل گرفته‌اند.

همچنین، انشعاب مفاهیم پایه به سمت حوزه‌های کاربردی نظیر «سرطان پستان»، «مدیریت ارتباط با مشتری» و «یادگیری ماشین» در بازه اخیر، از منظر مدل کریسپ تفسیرپذیر است: محققان نه‌تنها با الگوریتم‌ها آشنا شده‌اند، بلکه چرخه حیات تحلیل داده را در بستر مسائل واقعی به کار می‌برند (Shearer, 2000). این روند با یافته‌های پژوهش‌های بین‌المللی همسو است؛ بوزتاش و همکاران (Boztas et al., 2024) در تحلیل داده‌کاوی آموزشی نیز نشان دادند که گذار از روش‌شناسی محض به کاربردهای تخصصی، مشخصه دوران بلوغ یک حوزه علمی است. با این حال، در پژوهش‌های فارسی، این انتقال هنوز در مرحله شکل‌گیری است و نشانه‌هایی از تمرکز بیش از حد بر معدود روش‌های مسلط دیده می‌شود که ممکن است تنوع روش‌شناختی را محدود کند.

شبکه هم‌رخدادی کلیدواژه‌ها نیز این ساختار را از زاویه‌ای دیگر نشان می‌دهد. شکل‌گیری ۱۰ خوشه مفهومی در شبکه مورد بررسی بیانگر آن است که پژوهش‌های فارسی داده‌کاوی حول یک موضوع واحد متمرکز نبوده‌اند، بلکه مجموعه‌ای از مفاهیم روش‌شناختی در کنار موضوعات و حوزه‌های کاربردی مختلف در شبکه قرار گرفته‌اند. در این ساختار، حضور هم‌زمان روش‌هایی مانند «درخت تصمیم»، «خوشه‌بندی» و «پیش‌بینی» در کنار موضوعات کاربردی نشان می‌دهد که ساختار موضوعی پژوهش‌های مورد بررسی ماهیتی چندمحوری دارد؛ به این معنا که روش‌های داده‌کاوی در ارتباط با مسائل و حوزه‌های کاربردی متنوع ظاهر شده‌اند. بنابراین، خوشه‌بندی موضوعات صرفاً فهرستی از حوزه‌های پژوهشی نیست، بلکه نشان می‌دهد مفاهیم روش‌شناختی و کاربردی چگونه در ساختار موضوعی مجموعه مورد مطالعه در کنار یکدیگر قرار گرفته‌اند.

نتایج نمودار راهبردی، این یافته‌ها را با تمرکز بر جایگاه و ساختار ارتباطی خوشه‌ها تکمیل می‌کند. نبود خوشه‌ای در ربع مضامین محرک نشان می‌دهد که در شبکه حاضر خوشه‌ای با مرکزیت و تراکم هم‌زمان بالا مشاهده نشده است. با توجه به آستانه حداقل سه رخداد، ورود ۶۳ کلیدواژه به شبکه و ویژگی‌های پراکندگی داده‌ها، این نتیجه باید به ساختار شبکه مورد بررسی محدود شود و نمی‌توان آن را به‌تنهایی نشانه عدم بلوغ حوزه داده‌کاوی در ایران دانست. در مقابل، قرار گرفتن «درخت تصمیم»، «شبکه عصبی مصنوعی» و «ماشین بردار پشتیبان» در بخش مضامین پایه نشان می‌دهد که این مفاهیم در شبکه حاضر با موضوعات متنوع‌تری ارتباط دارند، در حالی که انسجام درونی آن‌ها نسبتاً پایین‌تر است. از سوی دیگر، قرار گرفتن «کیفیت آب» و «تحلیل مکانی» در بخش موضوعات

تخصصی نشان‌دهنده تراکم بیشتر و مرکزیت کمتر این موضوعات در شبکه است. این تفاوت در جایگاه خوشه‌ها نشان می‌دهد که موضوعات موجود در پژوهش‌های فارسی داده‌کاوی از نظر نحوه اتصال به ساختار کلی شبکه و میزان انسجام درونی وضعیت یکسانی ندارند.

قرار گرفتن «شبکه اجتماعی» و «اینستاگرام» در ربع پایین سمت چپ نیز نشان می‌دهد که این موضوعات در شبکه مورد بررسی از مرکزیت و تراکم پایین‌تری برخوردارند. با این حال، این جایگاه به‌تنهایی جهت زمانی موضوعات را مشخص نمی‌کند و نمی‌توان صرفاً بر اساس نمودار راهبردی آن‌ها را نوظهور یا رو به زوال دانست. این نکته اهمیت استفاده هم‌زمان از چند سطح تحلیل را نشان می‌دهد؛ زیرا تحلیل زمانی برای بررسی تغییر حضور موضوعات، تحلیل تکامل موضوعی برای بررسی روابط و مسیرهای ارتباطی، شبکه هم‌رخدادی برای شناسایی ساختار خوشه‌ها و نمودار راهبردی برای بررسی جایگاه و نقش ارتباطی خوشه‌ها به کار گرفته شده‌اند و هیچ‌یک به‌تنهایی تصویر کاملی از ساختار موضوعی ارائه نمی‌کنند.

از منظر کلی، کنار هم قرار دادن نتایج چهار تحلیل نشان می‌دهد که ساختار پژوهش‌های فارسی داده‌کاوی در مجموعه مورد مطالعه را می‌توان با سه ویژگی اصلی توصیف کرد: تداوم یک هسته از مفاهیم روش‌شناختی، تغییر تدریجی و بازترکیب واژگان و روابط موضوعی، و هم‌زیستی موضوعات پایه با حوزه‌های کاربردی و تخصصی. تداوم کلیدواژه‌هایی مانند «درخت تصمیم»، «شبکه عصبی مصنوعی» و «خوشه‌بندی» نشان‌دهنده ثبات نسبی بخشی از هسته روش‌شناختی است؛ در حالی که افزایش حضور برخی اصطلاحات در دوره‌های پایانی و تغییر ارتباط میان مفاهیم روش‌شناختی و کاربردی، پویایی ساختار موضوعی را نشان می‌دهد. از سوی دیگر، تفاوت میان جایگاه مضامین پایه و موضوعات تخصصی در نمودار راهبردی نشان می‌دهد که این مفاهیم از نظر میزان ارتباط با ساختار کلی شبکه و انسجام درونی، نقش یکسانی ندارند.

در انتها، نکته ای که باید بدان توجه داشت این است که نتایج پژوهش حاضر به جامعه مقالات فارسی نمایه‌شده در آی‌اس‌سی محدود است و بنابراین نباید به کل ادبیات داده‌کاوی ایران تعمیم داده شود. همچنین، کلیدواژه‌ها نماینده موضوعات مورد اشاره در مقالات هستند و لزوماً تمام جزئیات روش‌شناختی یا محتوای واقعی پژوهش‌ها را منعکس نمی‌کنند. در نتیجه، برای بررسی دقیق‌تر تغییرات روش‌شناختی و موضوعی این حوزه، پیشنهاد می‌شود در پژوهش‌های آینده علاوه بر کلیدواژه‌ها، متن کامل مقالات، چکیده‌ها، استنادها، همکاری‌های علمی و داده‌های مربوط به روش‌های واقعاً به‌کاررفته نیز بررسی شود و مقایسه با پایگاه‌های بین‌المللی و مجموعه‌های داده گسترده‌تر صورت گیرد.

پیشنهادهای اجرایی پژوهش

با توجه به یافته‌های پژوهش حاضر، پیشنهادهای اجرایی زیر ارائه می‌شود:

– سیاست‌گذاران و نهادهای تأمین‌کننده منابع پژوهشی می‌توانند از پژوهش‌هایی حمایت کنند که پیوند میان روش‌های داده‌کاوی و حوزه‌های کاربردی شناسایی‌شده در این مطالعه، مانند سلامت، مدیریت و تحلیل مکانی، را با طراحی‌های پژوهشی دقیق‌تر بررسی می‌کنند. همچنین، تقویت پژوهش‌هایی که از چند روش علم‌سنجی مکمل استفاده می‌کنند می‌تواند به تفسیر جامع‌تر ساختار موضوعی کمک کند.

–دانشگاه‌ها و مراکز پژوهشی، با ایجاد و تقویت گروه‌های پژوهشی میان‌رشته‌ای، زمینه تعامل بیشتر میان متخصصان علوم رایانه و پژوهشگران حوزه‌های کاربردی نظیر سلامت، مدیریت و علوم اجتماعی را فراهم سازند.

– برنامه‌ریزان آموزشی، می‌توانند با توجه به روند تغییر موضوعات و گسترش کاربردهای داده‌کاوی در حوزه‌های مختلف، سرفصل‌های درسی مرتبط را به گونه‌ای بازنگری کنند که علاوه بر مبانی الگوریتمی، کاربردهای میان‌رشته‌ای و مطالعات موردی نیز در آن‌ها مورد توجه قرار گیرد.

– نهادهای علمی و پژوهشی می‌توانند از اجرای پروژه‌های مشترک ملی و بین‌المللی در حوزه داده‌کاوی حمایت کنند تا امکان مقایسه الگوهای موضوعی پژوهش فارسی با ادبیات بین‌المللی و بررسی عوامل مؤثر بر انتخاب موضوعات و روش‌ها فراهم شود. این پیشنهاد با توجه به نتایج حاضر به‌عنوان یک مسیر تکمیلی برای پژوهش‌های آینده مطرح می‌شود، نه به‌عنوان راهکاری برای رفع یک علت اثبات‌شده در این مطالعه.

– برای بررسی اینکه آیا تفاوت در حضور روش‌های مختلف داده‌کاوی ناشی از دسترسی به داده، زیرساخت محاسباتی یا عوامل دیگری است، مطالعات آینده می‌توانند این متغیرها را به‌صورت مستقیم بررسی و با نتایج کتاب‌سنجی ترکیب کنند. داده‌های پژوهش حاضر به‌تنهایی امکان نسبت دادن الگوهای مشاهده‌شده به محدودیت زیرساختی را فراهم نمی‌کند.

–مدیران پایگاه‌های اطلاعات علمی، به‌ویژه مؤسسه استنادی و پایش علم و فناوری جهان اسلام (آی‌اس‌سی)، نسبت به توسعه قابلیت‌های نمایه‌سازی، استانداردسازی فراداده‌ها و ارائه خروجی‌های جامع استنادی اقدام کنند تا انجام مطالعات علم‌سنجی پیشرفته تسهیل شود.

پیشنهاد برای پژوهش های آتی

با توجه به یافته‌های این پژوهش و محدودیت‌های آن، پیشنهادهای زیر برای پژوهش‌های آتی ارائه می‌شود:

۱. مطالعه تطبیقی بین‌المللی: پژوهش‌های آتی می‌توانند با مقایسه همزمان تولیدات علمی پژوهشگران ایرانی در پایگاه‌های آی‌اس‌سی، اسکوپوس و وب‌آوساینس، تصویر جامع‌تری از جایگاه علم داده‌کاوی ایران در سطح منطقه‌ای و بین‌المللی فراهم کنند. چنین مطالعه‌ای امکان شناسایی شکاف‌های واقعی میان پژوهش‌های داخلی و بین‌المللی را فراهم می‌آورد.

۲. تحلیل‌های استنادی و هم‌استنادی: توسعه زیرساخت‌های استنادی در پایگاه آی‌اس‌سی می‌تواند زمینه تحلیل‌های هم‌استنادی را فراهم کند که تکمیل‌کننده نتایج این پژوهش خواهد بود و می‌تواند تصویر دقیق‌تری از وضعیت علمی در حوزه داده‌کاوی فارسی ارائه دهد.

۳. تحلیل شبکه هم‌نویسندگی و همکاری سازمانی: پژوهش‌های آتی می توانند به‌طور اختصاصی الگوهای همکاری علمی میان نویسندگان، دانشگاه‌ها و پژوهشگاه‌های فعال در حوزه داده‌کاوی را بررسی کنند. شناسایی عوامل تسهیل‌کننده یا بازدارنده همکاری‌های بین‌دانشگاهی و بین‌رشته‌ای می‌تواند اطلاعات ارزشمندی برای سیاست‌گذاران علمی فراهم کند.

۴. تحلیل محتوایی کیفی: ترکیب رویکردهای کمّی این پژوهش با تحلیل محتوایی کیفی و مصاحبه با پژوهشگران برجسته حوزه داده‌کاوی، می‌تواند عوامل زمینه‌ای مؤثر بر شکل‌گیری و تحول موضوعات پژوهشی را روشن سازد. درک انگیزه‌های محققان برای انتخاب موضوع و روش، بُعد تفسیری مهمی به تحلیل‌های کمّی اضافه می‌کند.

بیانیه‌ها و شفافیت پژوهش

■ بیانیه‌ی تعارض منافع و اخلاق پژوهش

نویسندگان این اثر اعلام می‌دارند که هیچ‌گونه تضاد منافع مالی یا غیرمالی در فرایند پژوهش و انتشار این دست‌نوشته وجود ندارد. همچنین، نویسندگان متعهد می‌شوند که تمامی اصول اخلاق نشر- شامل اصالت اثر و پرهیز از سرقت ادبی، اخذ رضایت آگاهانه، پرهیز از ساخت و دستکاری داده‌ها (جعل و تحریف)، خودداری از ارسال موازی و انتشار چندباره، و پایبندی به دستورالعمل مجله پیرامون کاربرد فناوری‌های هوش مصنوعی- را به‌دقت رعایت کرده‌اند.

■ گرنت‌ها و حمایت مالی

این مطالعه حا صل پژوه شی م ستقل ا ست و هیچ‌گونه حمایت مالی یا گرنت پژوه شی از سازمان‌های دولتی، تجاری (خصوصی) یا بخش غیرانتفاعی دریافت نکرده است.

■ سهم/ مشارکت نویسندگان

۱. حامد برنگی [مفهوم‌سازی، روش‌شناسی، نرم‌افزار، بررسی و کاوش، مدیریت داده‌ها، نگارش– پیش‌نویس اولیه]؛ ۲. *علی‌اکبر خاصه* [مفهوم‌سازی، اعتبارسنجی، تحلیل رسمی، نگارش– بازبینی و ویرایش، تصویرسازی، سرپرستی، مدیریت پروژه].

■ دسترسی‌پذیری داده‌ها

داده‌های پشتیبان یافته‌های این پژوهش همزمان با ارسال نسخه اولیه دستنوشته در اختیار مجله قرار گرفته و هم اکنون نیز در مخزن عمومی Zenodo با شناسه دیجیتال پیوسته به نشانی https://doi.org/10.5281/zenodo.23837761 بارگذاری شده و به‌صورت آزاد در دسترس است.

■ تقدیر و تشکر

نویسندگان مراتب قدردانی صمیمانه‌ی خود را از سردبیر و داوران محترم و ناشناس «پژوهشنامه علم سنجی» به پاس ارائه‌ی نظرات، رهنمودها و بازخوردهای ارزنده‌ای که موجب ارتقای کیفیت علمی و ساختاری این مقاله شد، ابراز می‌دارند.

فهرست منابع

احمدی، ح.، و حسین پور، م. (۱۴۰۴). ترسیم ساختار مفهومی حوزهٔ دانش کارآفرینی در ایران: تحلیل هم واژگانی. *پژوهش نامه علم سنجی*، ۱۱(۱)، ۲۱۳-۲۳۴.
https://doi.org/10.22070/rsci.2024.18889.1721

حاجیان، ح.، و زرجینی، ا. (۱۴۰۲). تحلیلی بر کاربردهای داده کاوی در صنعت بیمه بر اساس شبکه هم رخدادی واژگان ها و شناسایی معتبرترین مجلات با شاخص استناد به پژوهش های علمی با استفاده از رویکرد علم سنجی. *پژوهشنامه بیمه*، ۱۳(۱)، ۷۱-۸۶. https://doi.org/10.22056/ijir.2024.01.06

جواهری، م.، وکیلی مفرد، ح.، امیری، م.، و خاصه، ع. (۱۴۰۰). ترسیم و تحلیل نقشه دانش حوزه پژوهش‌های زنان و زایمان با استفاده از تحلیل هم‌رخدادی واژگان. *پژوهش‌نامه علم‌سنجی*، ۷(۱۴)، ۱۳۷-۱۵۶.
https://doi.org/10.22070/rsci.2020.5289.1359

خاصه، ع.، بانگ آور، م.، قاضی زاده، ح.، و مختاری، ح. (۱۴۰۰). یک دهه فعالیت فصلنامه تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی (۱۳۸۷-۱۳۹۶): تحلیل کتاب‌سنجی و دیداری‌سازی برونداد علمی. *پژوهش‌نامه علم‌سنجی*، ۷ (۱۳)، ۵۳-۷۶.
https://doi.org/10.22070/rsci.2020.4724.1319

رفیعی خشنود، م.، و کرامت‌فر، ع. (۱۴۰۱). مقایسه تحلیلی مدارک علمی ایران در حوزه موضوعی متن‌کاوی. *مجله علم سنجی کاسپین*، ۹(۱)، ۱۰۴-۱۱۶. http://cjs.mubabol.ac.ir/article-1-210-fa.html

سهیلی، ف.، خاصه، ع.، و کرانیان، پ. (۱۳۹۷). روند موضوعی مفاهیم حوزه علم اطلاعات و دانش شناسی ایران بر اساس تحلیل هم رخدادی واژگان. *فصلنامه مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۲۹(۲)، ۱۷۱-۱۹۰.
https://doi.org/10.30484/nastinfo.2018.2233

عشقی، ف.، افضل طوسی، ع.، و حیدری، م. (۱۴۰۵). پویش‌های پژوهشی هنر محیطی در ایران: تحلیل هم‌واژگانی مقاله‌های فارسی‌زبان. *پژوهش نامه علم سنجی*، ۱۲(۱)، ۱۹۱-۲۱۸
https://doi.org/10.22070/rsci.2025.20380.1803

فرهنگی، س.، خاصه، ع.، و ابراهیمی دینانی، آ. (۱۳۹۷). ترسیم ساختار دانش در پژوهش‌های عرفانی: تحلیل هم‌واژگانی. *ادبیات عرفانی و اسطوره شناختی*، ۱۴(۵۱)، ۲۰۷-۲۳۹.

https://sanad.iau.ir/Journal/jmmlq/Article/1049904
نجار لشگری، س.، زارعی، ه.، خلخالی، ع.، و پالی، س. (۱۴۰۲). ترسیم نقشه ساختاری دانش مدیریت آموزشی در ایران: تحلیل هم رخدادی واژگان. *پژوهش نامه علم سنجی*، ۹(۱)، ۳۸۷-۴۰۸

Ahmadi, H., & Husseinpour, M. (2025). Mapping the conceptual structure of entrepreneurship knowledge in Iran: A co-word Analysis. *Scientometrics Research Journal*, *11*(1), 213-234. https://doi.org/10.22070/rsci.2024.18889.1721 **[In Persian]**.

Aimbetova, G. N., Bekniyazova, N. D., & Izemetov, S. B. (2025). Modern Trends in Intelligent Data Analysis. *Eurasian Journal of Academic Research*, *5*(5), 160-163. https://doi.org/10.5281/zenodo.15551017

Al-Ali, R., Mekimah, S., Zighed, R., Shishakly, R., Almaiah, M., Shehab, R., … Aldhyani, T. H. H. (2025). Evolution and gaps in data mining research: Identifying the bibliometric landscape of data mining in management. *Decision Science Letters*, *14*(2), 435–448. https://doi.org/10.5267/j.dsl.2024.12.011

Alam, S., Zardari, S., mehdi, U.- e-L., Abbas, M., Huda, N. U., & Khan, M. A. (2023). The Impact of Data Mining on Digital Libraries – A bibliometric Study. *VFAST Transactions on Software Engineering*, *11*(4), 124–137. https://doi.org/10.21015/vtse.v11i4.1700

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007

Aria, M., Cuccurullo, C., D'Aniello, L., & Spano, M. (2026). Biblioshiny and the SAAS Workflow: An integrated framework for transparent and reproducible science mapping. A demonstration through the replication of a study. *Journal of Informetrics*, *20*(3), 101837. https://doi.org/10.1016/j.joi.2026.101837

Baek, C., & Doleck, T. (2022). Educational data mining: A bibliometric analysis of an emerging field. *IEEE Access*, *10*. https://doi.org/10.1109/ACCESS.2022.3160457

Biblioshiny (2026). *Biblioshiny: The shiny app for bibliometrix*. https://www.bibliometrix.org/biblioshiny/

Boztaş, G. D., Berigel, M., & Altınay, F. (2024). A bibliometric analysis of Educational Data Mining studies in global perspective. *Education and Information Technologies*, *29*, 8961–8985. https://doi.org/10.1007/s10639-023-12170-0

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*(2), 191–235. https://doi.org/10.1177/053901883022002003

Chen, C. (2017). Science mapping: A systematic review of the literature. *Journal of Data and Information Science*, *2*(2), 1-40. https://doi.org/10.1515/jdis-2017-0006

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, *62*(7), 1382-1402. https://doi.org/10.1002/asi.21525

da Silva, J. B. N., Senna, P., Chousa, A., & Coelho, O. (2020). Data mining and operations research techniques in Supply Chain Risk Management: A bibliometric study. *Brazilian Journal of Operations & Production Management*, *17*(3), 1–14. https://doi.org/10.14488/BJOPM.2020.029

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*, 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070

dos Santos, B. S., Steiner, M. T. A., Fenerich, A. T., & Lima, R. H. P. (2019). Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018. *Computers & Industrial Engineering, 138*, 106120. https://doi.org/10.1016/j.cie.2019.106120

Eshghi, F., Afzaltousi, E. and Heydari, M. (2026). Research Dynamics of Environmental Art in Iran: A Co-word Analysis of Persian-Language Articles. *Scientometrics Research Journal*, *12*(1), 191-218. https://doi.org/10.22070/rsci.2025.20380.1803 **[In Persian]**.

Farhangi, S., Khasseh, A. A., Ebrahimi Dinani, A. (2018). Mapping Knowledge Structure in Mystical Researches: A Co-Word Analysis. Mytho-Mystic Literature Quarterly Journal, 14(51), 207-239. https://sanad.iau.ir/Journal/jmmlq/Article/1049904 **[In Persian]**.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17*(3), 37–54. https://doi.org/10.1609/aimag.v17i3.1230

Hajiyan, H., & Zarjini, A. (2023). A comprehensive analysis of keywords co-occurrence network and the most cited journals on data mining techniques in insurance industry using scientometrics approach. *Iranian Journal of Insurance Research*, *13*(1), 71-86. https://doi.org/10.22056/ijir.2024.01.06 **[In Persian]**.

Ho, Y.S., & Lwesya, F. (2025), Sustainable Tech's global ascent: An analysis of emerging eco-friendly innovations. *Development and Sustainability in Economics and Finance*, *7*, Article Number: 100091. https://doi.org/10.1016/j.dsef.2025.100091

Hu, Y., Yu, Z., Cheng, X., Luo, Y., & Wen, C. (2020). A bibliometric analysis and visualization of medical data mining research. *Medicine*, *99*(22), e20338. https://doi.org/10.1097/MD.0000000000020338

Javaheri,M. , Vakilimofrad,H. , Amiri,M. and Khasseh,A. A. (2021). Mapping Knowledge Structure of Obstetrics and Gynecology studies: A Co-Word Analysis. Scientometrics Research Journal, 7((Issue 2, Autumn & Winter)), 137-156. https://doi.org/10.22070/rsci.2020.5289.1359 **[In Persian]**.

Khasseh,A. A. , BangAvar Mashak,M. , Ghazizadeh,H. and Mokhtari,H. (2021). A Decade of Performance of the Research on Information Science and Public Libraries Quarterly (2008-2017): A Bibliometric Analysis and Visualization of Scientific Output. Scientometrics Research Journal, 7((Issue 1, spring & summer)), 53-76. https://dor.isc.ac/dor/20.1001.1.24233773.1400.7.13.4.9 **[In Persian]**.

Khasseh, A. A., Soheili, F., Moghaddam, H. S., & Chelak, A. M. (2017). Intellectual structure of knowledge in iMetrics: A co-word analysis. *Information Processing & Management*, *53*(3), 705-720. https://doi.org/10.1016/j.ipm.2017.02.001

Kolling, M. L., Furstenau, L. B., Sott, M. K., Rabaioli, B., Ulmi, P. H., Bragazzi, N. L., & Tedesco, L. P. C. (2021). Data Mining in Healthcare: Applying Strategic Intelligence Techniques to Depict 25 Years of Research Development. *International Journal of Environmental Research and Public Health*, *18*(6), 3099. https://doi.org/10.3390/ijerph18063099

Kousis, A., & Tjortjis, C. (2021). Data mining algorithms for smart cities: A bibliometric analysis. *Algorithms*, *14*(8), 242. https://doi.org/10.3390/a14080242

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Leydesdorff, L. (2001). *The challenge of Scientometrics: The Development, Measurement, and Self-Organization of Scientific Communications* (2nd ed.). Universal Publishers. https://www.amazon.com/Challenge-Scientometrics-Development-Self-Organization-Communications/dp/1581126816

Mokhtarpour, R., & Khasseh, A. A. (2020). Who is who in library and information science research? The integrative application of scholarly influence indicators. *Journal of Librarianship and Information Science, 52*(4), 1186-1196. https://doi.org/10.1177/0961000620907956

Nagendhra, R. Y., & Jen, C. C. (2024). Bibliometric insights into data mining in education research: A decade in review. *Contem-porary Educational Technology, 16*(2). https://doi.org/10.30935/cedtech/14333

Najjar Lashgari, S., Zarei, H., Khalkhali, A. and Pali, S. (2023). Mapping the Intellectual Structure in the Field of Educational Management in Iran: Co-Word Analysis. *Scientometrics Research Journal*, *9*(1), 387-408. https://doi.org/10.22070/rsci.2021.14863.1517 **[In Persian]**.

Pynadath, M. F., Rofin, T. M., & Thomas, S. (2022). Evolution of customer relationship management to data mining-based customer relationship management: a scientometric analysis. *Quality & Quantity*, 1–32. https://doi.org/10.1007/s11135-022-01500-y

Rafiee, M., & Keramatfar, A. (2022). Analytical Comparison of Iranian Scientific Documents in Text Mining. *CJS, 9*(1), 104-116. http://cjs.mubabol.ac.ir/article-1-210-fa.html **[In Persian]**.

Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, *2*(5), 377. https://doi.org/10.1007/s42979-021-00765-8

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing, 5*(4), 13–22.

Soheili, F. , Khasseh, A. and koranian, P. (2018). Thematic trends of concepts in Knowledge and Information Science based on co-word analysis in Iran. *Librarianship and Information Organization Studies*, *29*(2), 171-190. https://doi.org/10.30484/nastinfo.2018.2233 **[In Persian]**.

Tseng, S. F., Won, Y. L., & Yang, J. M. (2016). A bibliometric analysis on Data Mining and Big Data. *International Journal of Electronic Business*, *13*(1), 38–69. https://doi.org/10.1504/IJEB.2016.075333

Upadhyay, A. (2018). A review on bibliometric analysis of data mining. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), 5*(4), 621-625. https://www.technoarete.org/common_abstract/pdf/IJERCSE/v5/i4/Ext_25017.pdf

Verma, R., Kumar, A., Pabreja, K., et al. (2025). A decade of data science research: Insights from a bibliometric study. *Discover Data*, *3*, 61. https://doi.org/10.1007/s44248-025-00097-5

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2011). *Data mining: Practical machine learning tools and techniques* (3th ed.). Morgan Kaufmann. https://doi.org/10.1016/C2009-0-19715-5

Declarations and Research Transparency

■ Conflict of Interest and Ethical Statement

The authors declare no conflict of interest regarding the publication of this manuscript. Furthermore, all ethical standards, including plagiarism, informed consent, misconduct, data fabrication/falsification, duplicate publication, redundant submission, and the journal's policy on the use of artificial intelligence (AI), have been strictly observed.

■ Funding & Financial Support

This study was conducted as an independent research project and received no specific grant or financial support from any funding agency in the public, commercial, or not-for-profit sectors.

■ Authors' Contributions/ CRediT Statement

HB: [Conceptualization, Methodology, Software, Investigation, Data Curation, Writing– original draft]; AAK: [Conceptualization, Validation, Formal Analysis, Writing – Review & Editing, Visualization, Supervision, Project Administration].

■ Data Availability Statement

The supporting data for the findings of this research were provided to the journal at the time of initial manuscript submission and are now also deposited in the public repository Zenodo under the persistent digital identifier at the following address [https://doi.org/10.5281/zenodo.23837761], where they are freely available.

■ Acknowledgments

The authors would like to express their sincere gratitude to the Editor-in-Chief and the anonymous peer reviewers of the *Scientometrics Research Journal* for their valuable feedback and constructive comments, which substantially improved the quality of this manuscript.

زودآیند ویرایش نشده