Letter to Editor

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2023; 10(2): 198-201

Knowledge Discovery and Data Mining in Medical Sciences, a Review of the Requirements for Applied Researches

Sarafi Nejad Afshin[1]*

• **Citation:** Sarafi Nejad A. Knowledge Discovery and Data Mining in Medical Sciences, a Review of the Requirements for Applied Researches. Journal of Health and Biomedical Informatics 2023; 10(2): 198-201. [In Persian]

1. Clinical Informatics Research and Development Lab, Clinical Research Development Unit, Shafa Hospital, Kerman University of Medical Sciences, Kerman, Iran
***Corresponding Author:** Afshin Sarafi Nejad
Address: University of Medical Sciences, Shafa Medical Education Center, Kausar Blvd., Kerman, Iran
• **Tel:** 09131414405 • **Email:** asarafinejad@kmu.ac.ir

© 2023 The Author(s); Published by Kerman University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cite

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.25]

نامه به سردبیر

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره دهم شماره دوم. ص ۲۰۱-۱۹۸، ۱۴۰۲

اکتشاف دانش و داده‌کاوی در علوم پزشکی، مروری بر الزامات پژوهش‌های کاربردی

افشین صرّافی نژاد[1]*

• دریافت مقاله: ۱۴۰۲/۵/۹ • پذیرش مقاله: ۱۴۰۲/۶/۱۴

نظم پیچیده و نهفته در اعماق داده‌های حجیم، همچون یک هزارتوی چندبعدی، توسط ذهن و زمان محدود انسان به آسانی قابل درک نیست. حل مسئله‌هایی با ابعاد فراوان و در تعداد بی‌شمار، بدون ابزار، بسیار دشوار و چه بسا غیرممکن است. «شیء که زمین نامیده شده، گرد است». این جمله که یک حقیقت بدیهی است، توسط گروهی از پژوهشگران آمریکایی با پردازش قوانین حاصل از مجموعه حجیم داده‌های گردآوری شده توسط انسان تا قرن دوم میلادی از خورشید، گردش زمین و شب و روز، در دهه هفتم قرن بیستم به دست آمد [۱]. در عصر اطلاعات، با پردازش کامپیوتری روابط بین صفات و داده‌های گوناگون، به مراتب سریع‌تر از اکتشاف علوم تجربی و شناخت عقلی، به کشف قوانین دانشی حیرت‌انگیز می‌رسیم. با مرور ساده‌ای بر تعریف استاندارد اکتشاف دانش و داده‌کاوی، این هدف را دنبال می‌کنیم که محققین جوان و علاقه‌مند به علوم داده با گرایش فنی مهندسی یا حوزه سلامت، به چه الزامات و نکات کلیدی مربوط به دانش داده کاوی تخصصی علوم پزشکی، باید توجه دقیق‌تری داشته باشند. در تصمیم‌گیری برای اجرای پروژه‌های داده کاوی در زمینه‌های تخصصی، نخست درک مفاهیم اصلی و کلیات و سپس تلاش برای شناخت جزئیات تخصصی نهفته در اعماق کتب و مقالات، رویکرد متفاوتی به ذهن کاوشگر می‌دهد.

اکتشاف دانش یا Knowledge Discovery و داده کاوی یا Data Mining دو کلیدواژه تخصصی هستند که پیشینه‌ای نسبتاً قدیمی دارند، اما از حدود دو دهه قبل به عنوان یک مسیر حرکت جدی در آینده ای که تقریباً امروز در آن هستیم پیش بینی شده بود. با مرور منابع معتبر و پراستناد، ساده‌ترین تعریفی که از این دانش می توان بازگو کرد چنین است: «داده‌کاوی فرایند آشکار کردن الگوهای هدف و دانش از میان حجم عظیمی از داده‌ها است» [۴-۲]. در این تعریف به ظاهر ساده، بلکه خردمندانه و ظریف؛ چند کلیدواژه پرمعنا خودنمایی می‌کنند: فرایند، آشکارسازی، الگوها، هدف، دانش، حجم عظیم و داده‌ها. بیایید قدری بیشتر به این نکات بیاندیشیم.

- **فرآیند:** همواره شروعی دارد، گام‌هایی به ترتیب و پیاپی برداشته می‌شود، با هدفی مشخص ادامه و پایان می‌یابد.
- **آشکارسازی:** دانسته‌ها را تکرار نکنیم. روابط پنهان بین متغیرها اغلب ناشناخته‌اند و وقتی عیان شوند، ارزشمندند. برای آشکارشدن زیبایی‌ها، باید به دنبال حقایق بود و اضافات را زدود. پاکسازی و تجمیع داده‌ها و انتخاب و حذف و تبدیل آن‌ها اقداماتی برای آشکارسازی هستند.
- **الگوها:** مفهوم الگو یعنی دسته‌ای از موارد تکرارشونده یا پیاپی که اغلب روند تکرارپذیری آن‌ها، یا به دنبال هم آمدن آن‌ها، در بین یک جمعیت بزرگ، ویژگی خاصی را در نهایت به همراه دارد. اگر این ویژگی، سودمند و به سادگی قابل درک باشد، می‌تواند یک الگوی جالب تلقی شود [۳،۴]. اکتشاف الگوهای جالب بین داده‌ها که پیشتر، نهفته بوده و یا با یافته‌های ذهنی و عینی، اما اثبات نشدهٔ قبلی تطبیق دارد، به هیجان این مسیر می‌افزاید. در این میان اصطلاحات تخصصی الگو و مدل، هر یک جایگاهی خاص دارند که نیاز به بحث مفصل تر دارد. [۱،۵].
- **هدف و دانش:** این دو همواره با هم هستنند، هر گاه به هدفی دست یابید، یعنی دانشی جدید را کشف کرده اید. بر خلاف پژوهش‌های مرسوم، در مسیر اکتشاف دانش، اهداف و دانش نهفته‌اند؛ کاوشگر نمی‌داند به چه دست خواهد یافت، بلکه فقط می‌کاود. حقیقت دیگر آن است که کاوشگر داده، داده

[Downloaded from jhbmi.com on 2024-10-12]
[DOI: 10.34172/jhbmi.2023.25]

را نمی‌کاود، بلکه دانش را بین انبوه داده‌ها می‌کاود و آن گاه که به آن دست می‌یابد، به هدف رسیده است [۴،۶].

- **حجم عظیم:** آسان نیست، اما الزامی است. وقتی سخن از حجم عظیم داده به میان می‌آید، به معنای میلیون‌ها ثبت داده‌ای و صفات مرتبط به هم است که ترکیب آن‌ها ممکن است نقشی کلیدی در شکل‌گیری یک پیامد داشته باشند، اما ارتباطات برخی از آن‌ها ناشناخته است [۷، ۸].
- **روش:** در تعریف به روش کشف دانش اشاره‌ای نشد، اما به مرور زمان الگوریتم‌ها و مدل‌های ریاضی متنوعی ساخته و پرداخته شده‌اند که اغلب در ابزارهای قدرتمند داده کاوی، با چند کلیک ساده در دسترس‌اند و سرعت عمل پژوهشگران را نیز بسیار بالا می‌برند [۵،۸].

در ایران، کاربرد داده‌کاوی در بخش خصوصی و با مدیریت گروه‌های فنی، به شکل جدّی و روزافزون متمرکز بر بازار و امور مالی است و کمتر به حل مسئله در حوزه علوم پزشکی پرداخته‌اند. صرف نظر از دلایل قابل بحث این عدم اقبال، نگاهی به مقالات پژوهشی و گزارش‌های پایان نامه‌ای تحصیلات تکمیلی علوم پزشکی ایران در سطح ملی و بین‌المللی نکاتی را می‌نمایاند. در بازه ۱۵ سال اخیر، نزدیک به ۷۰۰ طرح و پایان‌نامه مقاطع کارشناسی ارشد و دکترای تخصصی و بالغ بر ۳۰۰۰ مقاله علمی پژوهشی به زبان فارسی و در مجلات داخلی در زمینه اکتشاف دانش و داده‌کاوی در علوم پزشکی، منتشر شده و برخی مؤسسات خصوصی نیز تسهیلات و تلاش‌هایی برای توسعه این امر فراهم و دنبال کرده‌اند. این آمار که کمتر از یک درصد کل مجموعه پژوهش‌های علوم پزشکی کشور، به‌جزء مقالات غیرفارسی است، نشان می‌دهد نه تنها هنوز در زمینه داده کاوی حرف زیادی گفته نشده، بلکه مسیری با دشواری‌های خودش، باز و پیش رو است.

با نگاهی مروری و فارغ از نقد جدی مقالات تکنیکال داده‌کاوی که توسط محققین ایران در مجلات معتبر فارسی یا بین‌المللی منتشر گردیده، دیده می‌شود که به دو مسئله خاص، قدری کم توجهی شده است. موضوع اول، تمرکز بر حجم عظیم داده‌ها است. وقتی صحبت از حجم عظیم می‌شود، منظور چند صد یا چند هزار رکورد نیست، بلکه کاوشگر به دنبال دانش جدید و جالبِ نهفته در بین میلیون‌ها رکورد داده‌ای متنوع و در هم ریخته است که علی‌رغم آشفتگی ظاهری، تکرار، نقص، اشتباه یا عدم وجود قالب صحیح و حتی فقدان بخش‌های بزرگی از داده‌ها، باز هم به دلیل خصلت ذاتی انبوه بودنشان، رگه‌هایی از الگوهای جالب نیز در بین آن‌ها است که دست توانا و چشم تیزبین متخصص یا صاحب‌نظر آن حوزه دانش، با آگاهی درونی از آنچه که می‌جوید، می‌تواند ارزشمندی آن رگه را کشف کند و تلاشش به هدف برسد. موضوع دوم و مهم‌تر، اهمیت مشارکت فکری و نقش کلیدی متخصصین حوزه دانشی مورد نظر به‌ویژه در تحلیل و ترجمان دانش کشف شده می‌باشد. در شماری از مقالات، عدم ارتباط رشته تخصصی نویسندگان با موضوع اصلی پژوهش، شگفتی‌آور است. باید آگاهانه توجه داشت که یک گروه فنی متبحر و مسلط به استفاده از ابزارهای تکنیکی، زمانی موفق می‌شود که محصولش با نظر دقیق، و فراتر از آن، با مشارکت و نگاه تخصصی افراد دانش آموخته در زمینه مورد مطالعه، به درستی و به صورت کاربردی تحلیل و ارزیابی شود. در بیانی ساده، همان‌طور که متخصص علوم پزشکی نباید و نمی‌تواند در علوم مهندسی اظهارنظر تخصصی کند، بدون بهره‌گیری از مشورت متخصصین بالینی نیز بایسته نیست که گزارش‌ها صرفاً از جنبه فنی، ارتباطات و قواعد انجمنی و مدل‌های پیشگویی حاصل از داده های پزشکی یا یک بیماری شناخته شده، تحلیل و منتشر شوند.

تکیه بر معلومات عمومی و مهارت‌های پایه در کاوش داده‌های آزمایشی یا در حجم اندک یا تکرار الگوهای مرسوم روی داده‌های تمیز و پاکسازی شده بدون پایش و ارزیابی تخصصی، با آنچه به عنوان اصول اکتشاف دانش و داده‌کاوی می‌شناسیم مغایرت دارد و با اعطای مجوز نشر این قبیل پروژه‌ها، چه بسا پاسخ علمی چندان مطلوبی حاصل نشده و یا در عمل به کار مخاطبین اصلی نیاید.

تعارض منافع

در مقاله حاضر تعارض منافع، محدودیت اخلاقی، و حمایت مالی وجود ندارد.

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.25]

•**ارجاع:** صرّافی نژاد افشین. اکتشاف دانش و داده‌کاوی در علوم پزشکی، مروری بر الزامات پژوهش‌های کاربردی. مجله انفورماتیک سلامت و زیست پزشکی ۱۴۰۲؛ ۱۰(۲): ۲۰۰-۱۹۸.

۱. دفتر تحقیق و توسعه انفورماتیک بالینی، واحد توسعه تحقیقات بالینی، بیمارستان شفا، دانشگاه علوم پزشکی کرمان، کرمان، ایران

* **نویسنده مسئول:** افشین صرّافی نژاد

آدرس: کرمان، بلوار کوثر، مرکز آموزشی درمانی شفا، دانشگاه علوم پزشکی کرمان، کرمان، ایران. کد پستی: ۷۶۱۸۷۵۱۱۵۱

•**شماره تماس:** ۰۹۱۳۱۴۱۴۴۰۵ • **Email:** asarafinejad@kmu.ac.ir

References

1. Saniee Abadeh MS, Mahmoudi M. Taherparvar M. Applied Data Mining. 2nd ed. Tehran: Niaz Danesh; 2017. p. 536. [In Persian]

2. Han J, Kamber M, Pei J. Data Mining Concepts and Techniques. 3th ed. Morgan Kaufmann: Boston: 2012. p. 1-38.

3. Jothi N, Husain W. Data mining in healthcare–a review. Procedia Computer Science 2015;72:306-13. https://doi.org/10.1016/j.procs.2015.12.145

4. Fayyad UM, Piatetsky-Shapiro G, Smyth P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining; 1996 Aug 2-4; Portland Oregon: AAAI Press; 1996. p. 82-8.

5. Cios KJ, Pedrycz W, Swiniarski RW. Data mining methods for knowledge discovery. Springer Science & Business Media; 2012.

6. Kantardzic M. Data Mining: Concepts, models, methods, and algorithms. Technometrics. 2003;45(3):277.

7. Cummins MR, Nachimuthu SK, Abdelrahman SE, Facelli JC, Gouripeddi R. Nonhypothesis-Driven Research: Data Mining and Knowledge Discovery. InClinical Research Informatics. Cham: Springer International Publishing; 2023. p. 413-32.

8. Ganti V, Gehrke J, Ramakrishnan R. Mining very large databases. Computer 1999;32(8):38-45.

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.25]