Journal of Knowledge Studies
(Library and Information Science and Information Technology)
I. A. U. North Tehran Branch
Print ISSN: 2008-2754 Online ISSN: 2783-4115

Identifying Appropriate Data Attributes In Order to Knowledge Discovery from Health Big Data Systems

Fatemeh Soleimani-Roozbahani[1] | **Ali Rajabzadeh Ghatari**[2*] | **Reza Radfar**[3]

1-Ph.D. Candidate, Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran fatemeh.solaymani@srbiau.ac.ir
2-Department of Management, Tarbiat Modares University, Tehran, Iran (Corresponding Author) alirajabzadeh@modares.ac.ir
3-Department of Industrial Management, Science and Research Branch, Islamic Azad University, Tehran, Iran radfar@gmail.com

Article Info

Article type:
Research Article

Article history:
Received:
29 January 2022
Received in revised form:
13 March 2022
Accepted:
28 April 2022
Published online:
28 April 2022

Keywords:
Knowledge Discovery in Databases,
Health Big Data Systems,
Appropriate Data for Knowledge Discovery,
Exploratory and Confirmatory Factor Analysis

ABSTRACT

Objective: This study aimed to introduce the factors confirming the desirability of the data attributes in order to knowledge discovery from health big data systems by using factor analysis.

Methodology: This research was a Survey-Analytical study. The statistical population consisted of 655 specialists in the fields of data science, bioinformatics, and health information analysis inside and outside Iran, who were identified by snow flack method, and the research tool, researcher-made questionnaire was sent for them, and 247 questionnaires were completed and turned back. Data analysis methods were exploratory and confirmatory factor analysis using SPSS 22/0 and Lisrel 8/8 software.

Results: The results of Exploratory Factor Analysis explained six factors for identifying data characteristics that are: data quality, data stability, data compatibility, data value and accessibility, data efficiency and accuracy and completeness, which altogether had 68% of the total variance of the data. Confirmation analysis fit indices showed that the six-factor model has an acceptable fit with the data.

Conclusion: Findings indicate that based on the proposed model, data properties affect the process of knowledge discovery of health big data systems and six identified factors and 20 items of these factors, can be a good tool to identify the characteristics of data suitable for knowledge discovery of health big data systems for the effective use of these systems, emphasizing the independence of data format, and can help in the early diagnosis and treatment of diseases.

Cite this article: Soleimani-Roozbahani, F., Rajabzadeh Ghatari, A., Radfar, R. (2022). Identifying Appropriate Data Attributes In Order to Knowledge Discovery from Health Big Data Systems. *Journal of Knowledge Studies*, 15(56), 32-45.
DOR: 20.1001.1.20082754.1401.15.56.1.5

© The Author(s). **Publisher:** Islamic Azad University North Tehran Branch

دانش شناسی
(علوم کتابداری و اطلاع رسانی و فناوری اطلاعات)
شاپا الکترونیکی: ۴۱۱۵-۲۷۸۳ شاپا چاپی: ۲۷۵۴-۲۰۰۸

شناسایی ویژگی‌های داده درخور جهت کشف دانش از سیستم‌های کلان داده سلامت

فاطمه سلیمانی روزبهانی[۱] | **علی رجب زاده قطری**[۲*] | **رضا رادفر**[۳]

۱- دانشجوی دکتری رشته مدیریت فناوری اطلاعات/ کسب و کار هوشمند، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران fatemeh.solaymani@srbiau.ac.ir

۲- دانشیار گروه مدیریت صنعتی، دانشگاه تربیت مدرس، تهران ایران (نویسنده مسئول) alirajabzadeh@modares.ac.ir

۳- استاد گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران radfar@gmail.com

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۰/۱۱/۰۹
تاریخ بازنگری: ۱۴۰۰/۱۲/۲۲
تاریخ پذیرش: ۱۴۰۱/۰۲/۰۸
تاریخ انتشار آنلاین: ۱۴۰۱/۰۲/۰۸

واژه‌های کلیدی:
کشف دانش در دادگان، سیستم‌های کلان داده سلامت، داده‌های درخور جهت کشف دانش، تحلیل عاملی و تأییدی

چکیده

هدف: هدف مطالعه حاضر ارائه و معرفی عوامل تأیید کننده مطلوبیت ویژگی‌های داده‌های مورد استفاده جهت کشف دانش از سیستم‌های کلان داده سلامت بوده است.

روش پژوهش: پژوهش پیش‌رو مطالعه توصیفی-تحلیلی است. جامعه آماری متخصصان حوزه علوم داده، بیوانفورماتیک و تحلیل اطلاعات سلامت داخل و خارج از کشور بودند که با استفاده از روش گلوله برفی ۶۵۵ نفر از آنها شناسایی شده، و برای آنها ابزار پژوهش پرسشنامه محقق ساخته ارسال شد که ۲۴۷ پرسشنامه را تکمیل و برگرداندند. روش استفاده شده تحلیل داده‌ها، تحلیل عاملی اکتشافی و تأییدی با استفاده از نرم‌افزارهای SPSS ۲۶/۰ و Lisrel ۸/۸ بوده است.

یافته‌ها: نتایج تحلیل عاملی اکتشافی، شش عامل برای شناسایی ویژگی‌های داده درخور تبیین نمود که عبارتند از کیفیت داده، پایداری داده، سازگاری داده، ارزشمندی و دسترسی پذیری داده، کارایی داده و صحت و کامل بودن و روی هم رفته ۶۸ درصد از واریانس کل داده‌ها را دربرداشته است. شاخص‌های برازش تحلیل تأییدی نشان داد که الگوی شش عاملی برازش قابل قبولی با داده‌ها دارد.

نتیجه گیری: یافته‌های پژوهش حاکی از آن است که بر اساس مدل ارائه شده ویژگی‌های داده بر فرایند کشف دانش از سیستم‌های کلان داده سلامت تأثیر دارد و شش عامل شناسایی‌شده و ۲۰ گویه سازنده این عوامل، می‌تواند ابزار مناسبی جهت شناسایی ویژگی‌های داده‌های مناسب برای کشف دانش از سیستم‌های کلان داده سلامت جهت به‌کارگیری مؤثر این سیستم‌ها باشد که با تاکید بر عدم وابستگی به فرمت داده، خصایص فراگیر داده‌های حوزه سلامت را دربرگرفته و توجه به آن می‌تواند به تشخیص و درمان زودهنگام بیماری‌ها کمک شایانی کند.

استناد: سلیمانی روزبهانی، ف.، رجب زاده قطری، ع.، رادفر، ر. (۱۴۰۱). شناسایی ویژگی‌های داده درخور جهت کشف دانش از سیستم‌های کلان داده سلامت. *دانش شناسی*، ۱۵(۵۶)، ۳۲-۴۵.

DOR: 20.1001.1.20082754.1401.15.56.1.5

ناشر: دانشگاه آزاد اسلامی واحد تهران شمال

حق‌مؤلف © نویسندگان.

مقدمه

آنچه امروزه با عنوان کلان داده از آن یاد می‌شود مواد خام، دارایی‌های اطلاعاتی و دانشی سازمان‌ها است که در حجم بالا با سرعت زیاد تولید می‌شوند و از تنوع زیادی برخوردارند (میرحسینی و همکاران، ۱۴۰۰) و دانش حاصل از پردازش این کلان داده محور اصلی کسب‌وکار و علم در تمامی زمینه‌هاست (شوبانا و کومار[1]، ۲۰۱۵). خودکارسازی فرایندهای سازمان‌ها و بهبود خدمات‌دهی به مشتریان و تصمیم‌گیری سریع و دقیق سازمانی از مزایای بهره‌گیری از پردازش کلان داده است (میرحسینی و همکاران، ۱۴۰۰). حوزه سلامت نیز هم‌چون دیگر حوزه‌ها به شدت مبتنی بر داده است و قادر است از بسترهای کلان داده در کنار فناوری‌ها و ابزارهای نوآورانه برای پیشبرد اهداف مراقبتی و درمانی خود بهره گیرد (گالیتسی، کاتسالیاکی، و کومار[2]، ۲۰۲۰). کلان داده حوزه سلامت را می‌توان به چهار دسته اصلی طبقه‌بندی کرد: کلان داده پزشکی یا بالینی، کلان داده سلامت و رفتار عمومی، کلان داده تجارب پزشکی، و کلان داده محتوای پزشکی(هونگ و دیگران[3]، ۲۰۱۸). هر چهار حوزه کلان داده سلامت حاوی حجم عظیمی از داده‌های بی‌ساختار (مانند دست نوشته‌های کارکنان حوزه درمان) و داده‌های ساختاریافته متنوع، داده‌های حساس، داده‌های پردازش آنی است که ادغام، پردازش و ذخیره آنها را با چالش بزرگی روبرو می‌کند (جی و کیم[4]، ۲۰۱۳). با این حال مزایای بهره‌گیری از کلان داده پردازش شده حوزه سلامت و کشف دانش از آنها، انگیزه و اشتیاق کافی را برای ذینفعان این حوزه جهت بهره‌برداری از این منبع پیچیده فراهم کرده است (رادیک[5] و همکاران، ۲۰۰۳ و پرامانیک[6] و همکاران، ۲۰۲۲). بهره‌برداری از کلان داده حوزه سلامت مستلزم فراهم‌سازی داده مناسب و درخور استفاده در سیستم‌های پردازشی است که عدم توجه به آن هزینه قابل توجهی را با خود به همراه دارد. به طور مثال، کسب‌وکارهای ایالات متحده به طور متوسط سالانه بالغ بر ۶۰۰ میلیارد دلار در این حوزه هزینه می‌کنند (عبدالله[7]، ۲۰۱۵). مناسب و درخوربودن داده در برنامه‌های پردازش کلان داده که حجم، تنوع و سرعت بالایی از داده را پشتیبانی می‌کنند از اهمیت بالایی برخوردار است. داده نامناسب علاوه بر تأثیر منفی بر انتخاب و توسعه الگوریتم‌های پردازش، قادر است صدمات جبران‌ناپذیری در دنیای واقعی ایجاد کند (لوتون[8]، ۲۰۲۱). این مهم بر لزوم پرشدن شکاف میان داده‌های ذخیره شده در منابع داده و شکل مناسب آن جهت استفاده از الگوریتم‌های کشف دانش پیش از بهره‌گیری از ابزارهای کشف دانش صحه می‌گذارد (آلفرد[9]، ۲۰۰۵). مسئله مناسب و درخور بودن داده در حوزه سلامت از اهمیت، حساسیت، و عمق بیشتری برخوردار است. داده حوزه سلامت به سرعت در طول زمان به شیوه‌های مختلف همچون منسوخ شدن و تغییر، درخوربودن خود را از دست می‌دهد (عبدالله، ۲۰۱۵). این امر نه تنها فرایند پایش عملکرد مؤسسات و کارکنان حوزه درمان را تحت تأثیر قرار می‌دهد بلکه قادر است فرآیندهای پیشگیری، درمان و مراقبت را با چالش‌های جدی روبرو نماید و سلامت فردی و اجتماعی را به خطر اندازد. بر اساس اهمیت مسئله تناسب، هماهنگی و درخور بودن داده‌ها به منظور کشف دانش از آنها در سیستم‌های کلان داده سلامت، این پژوهش به دنبال آن است که ابعاد داده مناسب و درخور (دقت و اعتبار، قابلیت اطمینان، جامعیت، خوانایی، بهنگامی، قابلیت دسترسی، مفهوم یا مفید بودن، امنیت و محرمانه بودن) را از دیدگاه خبرگان حوزه علم داده و فناوری اطلاعات سلامت مورد ارزیابی قرار دهد تا در صورت امکان بتوان جهت ارتقاء سلامت فردی و اجتماعی و تسریع و بهبود تصمیمات و سیاست‌های کلان حوزه سلامت از دانش کشف شده از منبع داده‌ای عظیم سلامت بهره جوید. یافته‌های این پژوهش می‌تواند جهت افزایش نتیجه‌بخشی، کارایی و اثربخشی سیستم‌های ذخیره و پردازش کلان

1. Shobana and Kumar
2. Galetsi, Katsaliaki, and Kumar
3. Hong et al.
4. Jee and Kim 2013
5. Roddick
6. Pramanik
7. Abdullah
8. Lawton
9. Alfred

داده سلامت و در نتیجه آن بهبود سطح کیفی سلامت جامعه، سرعت بخشی به تصمیمات این حوزه و در نتیجه ارتقاء کیفیت زندگی شهروندان مثمرثمر باشد.

هرچند که داده درخور و مناسب تعریفی یکتا و واحد ندارد و بسته به محیط اطلاعاتی و نوع کاربرد می‌تواند شکل متفاوت به خود بگیرد، اما می‌توان بر اساس برآیند این پژوهش‌ها ویژگی‌های داده درخور را بازنمایی نمود. داده درخور داده‌ای است که دسترسی به آن همراه با سهولت و امنیت باشد، به عبارتی دسترسی‌پذیری یکی از ویژگی‌های داده مناسب و درخور است. قابلیت بالای درک و همچنین تفسیر داده، خوانایی، جامعیت و امکان خلاصه‌سازی و بازنمایی بصری، و ایجاد ارزش افزوده، مفید بودن آن، کافی و وافی بودن داده همچنین اعتبار، باورپذیری، صحت و دقت و اعتبار داده و بهنگامی آن و قابلیت یکپارچه‌سازی نیز از ویژگی‌های معرفی شده برای داده درخور استفاده در سیستم‌های پردازشی سلامت هستند (وانگ و استرانگ[1]، ۱۹۹۶، سازمان بهداشت جهانی، ۲۰۰۳ و مؤسسه اطلاعات سلامت کانادا[2]، ۲۰۱۱). پژوهش‌های انجام شده در این حوزه با رویکردهای مختلف مسئله مناسب و در خوربودن داده حوزه سلامت را مورد بررسی قرارداده‌اند. ذخیره، بازیابی و تبادل استاندارد و مناسب داده و اطلاعات در حوزه بهداشت و بخصوص در سیستم‌های پرونده الکترونیک سلامت از ضروریات استفاده از فناوری اطلاعات در این حوزه است و زبان مشترکی جهت توسعه سیستم‌های اطلاعات سلامت ایجاد می‌کند (احمدی و دیگران، ۱۳۹۱). اما از آنجا که الگوی جامع و استانداردی جهت توسعه پرونده الکترونیک سلامت وجود ندارد، برخی مطالعات مجموعه‌ای از استانداردهای توسعه پرونده الکترونیک سلامت را به منظور دسترسی به داده‌های مناسب، کامل و صحیح جهت تدوین استاندارد ملی پرونده الکترونیک سلامت بررسی نموده‌اند (کیخا و دیگران، ۱۳۹۵). از سویی ارائه خدمات مناسب و با کیفیت به بیماران از وظایف نظام‌های اطلاعات سلامت است که تأمین ویژگی‌های داده امن امکان بهره‌برداری کامل و موفق خدمات سلامت را فراهم می‌نماید (مقدسی و عیانی، ۱۳۹۲). بر همین اساس در راستای تأمین ویژگی‌های داده امن، استانداردهای بین المللی داده امن مناسب استفاده در سیستم‌های سلامت به ویژگی‌های داده و اطلاعات، آموزش، دسترسی و سرور شبکه توجه نموده‌اند (مقدسی و قائمی، ۱۳۹۴). پژوهشگرانی همچون واند[3]، وانگ، ردمن[4]، ژارک[5] و بووی[6] از جمله شاخص‌ترین نظریه‌پردازان این حوزه می‌باشند، که ویژگی‌های داده مناسب و درخور حوزه سلامت را قابلیت دسترسی، قابلیت بازنمایی، کیفیت زمینه‌ای و کیفیت ذاتی برشمرده‌اند. همچنین سازمان بهداشت جهانی ابعاد داده مناسب و درخور جهت استفاده از سیستم‌های الکترونیک سلامت را دقت، اعتبار قابلیت اطمینان، جامعیت، خوانایی، بهنگامی، قابلیت دسترسی، مفهوم یا مفید بودن، امنیت و محرمانه بودن برشمرده است (۲۰۰۳). در راستای به‌اشتراک‌گذاری این داده‌ها به شکل مناسب و مطلوب روشی جهت مناسب‌سازی داده و پاکسازی آن بمنظور منحصربه‌فرد شدن ویژگی‌های رکوردهای داده‌ای در مجموعه داده ارائه شده است (پرسر[7] و همکاران، ۲۰۱۸). با توجه به افزایش تمرکز و اهمیت گردآوری و پردازش داده‌های عظیم سلامت جهت مهار و مبارزه با پاندمی کووید ۱۹ ابزاری به منظور ارزیابی کیفیت و تناسب داده‌های حاکمیت باز کووید ۱۹ منتشر شده است که البته در حال حاضر داده‌های سلامت در ایالات متحده، اتحادیه اروپا و چین فاقد چنین ابرداده‌ای[8] است.

بررسی انتقادی و تحلیلی ادبیات حوزه سلامت، عدم وجود روش یا استاندارد مناسب شناسایی داده‌های درخور (مناسب) را جهت کشف دانش از کلان داده این حوزه نشان می‌دهد. در این راستا، پژوهش حاضر در نظر دارد ویژگی‌های داده درخور جهت کشف دانش از آنها در حوزه سلامت را شناسایی کند. جهت دستیابی به این هدف پژوهشگر از روش ترکیبی جمع‌آوری داده (مرور ادبیات حوزه و

---

1. Wang & Strong
2. Canadian Health Information Institute
3. Wand
4. Redman
5. Jarke
6. Bovee
7. Prasser
8. metadata

مصاحبه عمیق با خبرگان) بهره می‌جوید. ویژگی‌های استخراج و شناسایی شده با تاکید بر عدم وابستگی به فرمت داده، خصایصی فراگیر خواهد بود که داده‌های حوزه سلامت را دربرمی‌گیرد. با توجه به هدف پژوهش حاضر، سؤالات اصلی پژوهش شرح ذیل می‌باشند:

- ویژگی‌های داده درخور در حوزه سلامت جهت کشف بهینه دانش از آن‌ها کدامند؟
- آیا ویژگی‌های داده درخور بر فرایند کشف دانش از سیستم‌های کلان داده سلامت تأثیر دارد؟

روش پژوهش

پژوهش حاضر از نوع کاربردی و به روش کمی توصیفی- تحلیلی انجام شده است، همچنین از نظر هدف پژوهش، اکتشافی-تبیینی- پیش‌بینی است، و از نظر رویکرد پژوهش، کمی است. جامعه آماری، تمامی خبرگان در حوزه یادگیری ماشین و داده‌کاوی، علوم پزشکی، مدیریت اطلاعات و اسناد پزشکی، فناوری اطلاعات سلامت داخل و خارج از کشور هستند که تعداد دقیقی از آنها در دسترسی نیست. از آنجا که در تحلیل عاملی حجم نمونه برای هر متغیر پنج تا ۱۰ نمونه و طور کلی در مجموع تا حداکثر ۳۰۰ نمونه توصیه شده است، بر این اساس پرسشنامه برای ۶۵۵ نفر از خبرگان داخلی و خارجی شناسایی شده (بر اساس سوابق کاری و تحصیلی آنها در حوزه مورد نظر) با استفاده از روش گلوله برفی ارسال شد و ۲۴۷ پرسشنامه پاسخ داده شده برگشت داده شد (که نشان دهنده میزان پاسخگویی ۳۷/۷ درصد است).

ابزار استفاده شده برای گردآوری داده‌ها پرسشنامه محقق ساخته بر اساس بررسی و مرور متون و ادبیات موضوع است و دارای ۲۰ سؤال رتبه‌بندی شده بر اساس طیف لیکرت پنج گزینه‌ای که از خیلی کم تا خیلی زیاد نمره‌گذاری شده، بود. برای تأیید روایی محتوایی و صوری این پرسشنامه در اختیار ۲۰ نفر از اعضای هیأت علمی حوزه علم داده، مهندسی داده، تحلیل داده، انفورماتیک پزشکی و اطلاعات سلامت قرار داده شد و پرسشنامه نهایی بر اساس نظر آنها تدوین گردید. بر اساس نظر آنها پرسشنامه از روایی صوری و محتوایی مناسبی برخوردار بوده است. همچنین برای ارزیابی پایایی پرسشنامه در این پژوهش، ۲۰ نسخه به‌عنوان پیش‌آزمون[1]، جمع‌آوری‌شده و توسط نرم‌افزار SPSS ۲۶/۰[2]، آلفای کرونباخ برای آن‌ها محاسبه‌شده است که این آلفا ماهیت واریانس دارد. با انجام این آزمون روی داده‌ها آلفای کرونباخ برابر ۰/۹۲۲ به‌دست‌آمده است که نشان‌دهنده پایایی بالا[3] و در حد مطلوب مقیاس مورداستفاده برای بررسی موضوع مطالعه است.

در این پژوهش به منظور پی‌بردن به متغیرهای زیربنایی یک پدیده یا تلخیص داده‌ها از روش تحلیل عاملی استفاده شده است. داده‌های اولیه برای تحلیل عاملی، ماتریس همبستگی بین متغیرها است. تحلیل عاملی، متغیرهای وابسته از قبل تعیین شده‌ای ندارد. مقاصد استفاده از تحلیل عاملی به دو دسته کلی مقاصد اکتشافی، مقاصد تأییدی تقسیم می‌شود. در این پژوهش از هر دو مقصد تحلیل عاملی برای شناسایی و تأیید ویژگی‌های مناسب و درخور داده‌ها برای کشف دانش از سیستم‌های کلان داده سلامت استفاده شده است.

با استفاده از روش تحلیل عاملی اکتشافی توسط نرم‌افزار SPSS ۲۶/۰، داده‌های گردآوری شده تحلیل و بر اساس گام‌های این روش (ارزیابی تناسب داده‌ها برای تحلیل عاملی، استخراج عامل‌ها، چرخش عامل‌ها و نام‌گذاری عوامل) عوامل شناسایی داده در خور جهت کشف دانش شناسایی شد. سپس با استفاده از نرم‌افزار Lisrel ۸/۸ بر روی عوامل شناسایی شده، تحلیل عاملی تأییدی جهت تعیین میزان توان مدل عامل از قبل تعریف شده به شرح زیر انجام گردید.

به منظور شناسایی عوامل مؤثر برتأیید ویژگی‌های داده‌های مناسب جهت کشف دانش در سیستم‌های کلان داده سلامت از تحلیل عاملی اکتشافی متغیرها با استفاده از نظرات ۲۴۷ نفر از خبرگان استفاده شد. در این فرایند ابتدا کفایت نمونه‌گیری با استفاده از آزمون

---

1. Pilot Study
2. Statistical Package for Social Sciences 23.0

[3] درصورتی‌که ضریب آلفای کرونباخ پرسشنامه بزرگ‌تر از ۰.۷ باشد، آنگاه آزمون از پایایی قابل قبولی برخوردار است.

کرویت بارتلت[1] و آزمون کفایت نمونه‌گیری کیسر- مایر- الکین[2] بررسی شد. نتیجه آزمون KMO به همراه نتیجه آزمون بارتلت در جدول ۱ آمده است.

جدول ۱. نتایج آزمون KMO و Bartlett

KMO and Bartlett's Test		
آزمون کفایت نمونه‌گیری KMO		۰/۸۱
آزمون کرویت بارتلت	۳٥٤/۷۷۸	۷۷۸/۳٥٤
	۱۹۰	۱۹۰
	۰/۰۰۰	۰/۰۰۰

از آنجا که شاخص KMO برای این سازه برابر با ۰/۸۱ و شاخص بارتلت ۰/۰۰ است پس می‌توان نتیجه گرفت داده‌های جمع‌آوری شده برای این سازه به منظور استفاده در تحلیل عاملی مناسب هستند. گام بعدی فرایند، تعیین مقادیر اشتراکات استخراجی برای متغیرهاست. همبستگی هر متغیر با هر عامل بار عاملی نامیده می‌شود و مقدار آن بین ۱- و ۱+ تغییر می‌کند. در این مطالعه، مقادیر اشتراکات استخراجی برای متغیرها کوچکتر از یک بود. این عامل‌ها توانسته‌اند ۶۸ درصد از واریانس گویه شماره ۸ (اعتبار داده و قابل قبول بودن داده) و در مقابل ۳۸ درصد از واریانس گویه ۱۰ (مقبولیت منبع داده) را تبیین کنند. مرحله سوم فرایند تحلیل عاملی مشخص کردن سهم هر عامل در تبیین مجموع واریانس تمامی گویه‌ها است. هدف مرحله استخراج عامل‌ها، به دست آوردن سازه‌های زیربنایی است که تغییرات متغیرهای مورد مشاهده را موجب شده است. جدول ۲ عوامل استخراج شده همراه مقدار ویژه، درصد واریانس و درصد واریانس تجمعی را نشان می‌دهد.

جدول ۲. استخراج عامل‌های مؤثر بر شناسایی و تأیید ویژگی‌های داده

عامل‌ها	مقادیر ویژه اولیه			مقادیر ویژه عوامل استخراجی بدون چرخش			مقادیر ویژه عوامل استخراجی چرخش‌یافته		
	مجموع	درصد واریانس	درصد تجمعی	مجموع	درصد واریانس	درصد تجمعی	مجموع	درصد واریانس	درصد تجمعی
۱	۴/۲۸۰	۲۱/۳۹۹	۲۱/۳۹۹	۴/۲۸۰	۲۱/۳۹۹	۲۱/۳۹۹	۲/۱۲۸	۱۰/۶۴۰	۱۰/۶۴۰
۲	۱/۳۶۴	۶/۸۱۶	۲۸/۲۱۸	۱/۳۶۴	۶/۸۱۹	۲۸/۲۱۸	۱/۸۵۵	۹/۲۷۵	۱۹/۹۱۴
۳	۱/۳۳۰	۶/۶۵۲	۳۴/۸۷۰	۱/۳۳۰	۶/۶۵۲	۳۴/۸۷۰	۱/۸۰۸	۹/۰۴۱	۲۸/۹۵۵
۴	۱/۲۱۶	۶/۰۸۲	۴۰/۹۵۲	۱/۲۱۶	۶/۰۸۲	۴۰/۹۵۲	۱/۷۳۲	۸/۶۵۹	۳۷/۶۱۴
۵	۱/۱۶۰	۵/۷۹۹	۴۶/۷۵۱	۱/۱۶۰	۵/۷۹۹	۴۶/۷۵۱	۱/۴۴۵	۷/۲۲۶	۴۴/۸۴۱
۶	۱/۰۳۰	۵/۱۵۱	۵۱/۹۰۲	۱/۰۳۰	۵/۱۵۱	۵۱/۹۰۲	۱/۴۱۲	۷/۰۶۱	۵۱/۹۰۲
۷	۰/۹۹۷	۴/۹۸۳	۵۶/۸۸۶						
۸	۰/۹۱۴	۴/۵۶۸	۶۱/۴۵۳						

[1] آزمون کرویت بارتلت به ارزیابی این سؤال می‌پردازد: ماتریس همبستگی که پایه تحلیل عاملی قرار می‌گیرد، در جامعه برابر Baelett's Test of Sphericity صفر است یا خیر.

[2] مشخص می‌کند که آیا تحلیل عاملی بر روی داده‌های جمع آوری شده قابل اجرا می‌باشد KMO شاخص کایزر- مایر –الکین (KMO)

۹	۰/۸۶۹	۴/۳۴۴	۶۵/۷۹۸						
۱۰	۰/۸۵۴	۴/۲۷۰	۷۰/۰۶۷						
۱۱	۰/۷۶۹	۳/۸۴۴	۷۳/۹۱۱						
۱۲	۰/۷۴۵	۳/۷۲۵	۷۷/۶۳۶						
۱۳	۰/۶۸۳	۳/۴۱۴	۸۱/۰۵۰						
۱۴	۰/۶۴۴	۳/۲۱۸	۸۴/۲۶۸						
۱۵	۰/۶۳۳	۳/۱۶۴	۸۷/۴۳۱						
۱۶	۰/۵۷۶	۲/۸۷۸	۹۰/۳۰۹						
۱۷	۰/۵۴۷	۲/۷۳۶	۹۳/۰۴۶						
۱۸	۰/۵۱۱	۲/۵۵۵	۹۵/۶۰۱						
۱۹	۰/۴۶۱	۲/۳۰۷	۹۷/۹۰۸						
۲۰	۰/۴۱۸	۲/۰۹۲	۱۰۰/۰۰						

همانطور که در جدول ۲ دیده می‌شود ۲۰ گویه مورد نظر به ۶ عامل کاهش پیدا کرده است و می‌توان از ترکیب این ۲۰ گویه، ساختار جدیدی بر اساس عامل‌ها طراحی و بر اساس آنها به تحلیل پرداخت. و اگر شش عامل استخراج شود، حدود ۵۱ درصد از واریانس بیان خواهد شد. عامل اول با مقدار ویژه ۴/۲۸ حدود ۱۱ درصد از کل واریانس ویژگی‌های داده را تبیین نماید. عامل دوم با مقدار ویژه ۱/۳۶۴، بیش از ۹ درصد واریانس، عامل سوم با مقدار ویژه ۱/۳۳ نیز ۹ درصد از واریانس، عامل چهارم با مقدار ویژه ۱/۲۱۶، ۹ درصد از واریانس، عامل پنجم با مقدار ویژه ۱/۱۶، بیش از ۷ درصد واریانس و نهایتاً عامل ششم با مقدار ویژه ۱/۰۳ نیز ۷ درصد واریانس را تبیین نموده است.

مرحله چهارم از تحلیل عاملی، شناخت ماتریس همبستگی بین گویه‌ها و عامل‌ها و دسته‌بندی هر گویه در هر عامل است. در جدول ماتریس همبستگی بین گویه‌ها و عامل‌های استخراج شده با مقدار ویژه بالاتر از یک را قبل از چرخش که مقدارهمبستگی آن بین گویه‌ها و عامل‌ها بین ۱- و ۱+ نوسان است، معمولاً تفسیر نتایج همبستگی‌ها قبل از چرخش، با مشکل رو برو می‌شود، به همین دلیل نتایج قبل از چرخش در این مطالعه آورده نشده است.

جدول ۳ . ماتریس مولفه‌های تحلیل عاملی چرخش یافته برای شناسایی و تأیید ویژگی‌های داده

شماره گویه	متغیرها	عامل‌های چرخش یافته					
		۱	۲	۳	۴	۵	۶
q9	معقولیت داده /قابل قبول بودن داده	۰/۶۷۳	- ۰/۱۳۲	۰/۱۱۲	۰/۱۴۱	۰/۲۳۶	۰/۰۱۲
q18	دسترسی‌پذیری داده	۰/۶۴۷	۰/۱۹۷	- ۰/۰۱۴	۰/۲۹۱	-۰/۲۳۱	- ۰/۰۵۰
q8	اعتبار داده	۰/۵۷۴	۰/۱۸۲	۰/۰۵۹	-۰/۱۰۵	-۰/۲۱۹	۰/۴۷۵
q10	مقبولیت منبع داده	۰/۴۹۶	۰/۳۰۳	۰/۰۷۰	۰/۱۲۱	۰/۱۰۳	۰/۰۸۸
q5	قابلیت تفسیر داده	۰/۰۵۸	۰/۶۵۲	- ۰/۱۴۷	۰/۱۳۹	-۰/۰۱۵	۰/۳۱۱
q4	ابعاد مرتبط با زمان/ زمان محور بودن داده	۰/۰۴۲	۰/۵۸۶	۰/۲۴۱	۰/۱۶۹	۰/۰۶۰	۰/۱۲۶
q3	ثبات در سراسر منابع داده قابلیت اطمینان	۰/۳۸۸	۰/۵۷۵	۰/۱۵۵	-۰/۲۳۹	۰/۱۶۶	۰/۰۴۹
q20	پاسخگویی داده/ زمان پاسخ داده	۰/۰۵۹	۰/۵۷۱	۰/۲۱۱	۰/۰۹۱	۰/۲۲۸	- ۰/۰۷۴

q2	کامل بودن و جامعیت داده‌ها	۰/۲۵۴	۰/۰۶۵	۰/۶۳۳	۰/۰۱۹	-۰/۲۴۲	- ۰/۰۲۳
q11	عینی و مشهود بودن داده	۰/۰۰۸	۰/۱۰۰	۰/۶۳۲	۰/۱۴۲	۰/۲۲۷	۰/۱۹۱
q7	قابلیت اتکا داده	- ۰/۰۰۴	۰/۱۸۱	۰/۶۲۷	۰/۱۱۲	۰/۰۹۹	۰/۱۵۶
q17	حجم مناسب داده/ حجم داده	۰/۳۷۳	۰/۰۵۶	۰/۴۰۶	۰/۰۰۶	۰/۳۵۱	- ۰/۳۷۴
q19	قابلیت انتقال داده	۰/۰۵۶	۰/۰۷۹	۰/۰۰۱	۰/۷۱۹	۰/۱۵۴	۰/۱۴۲
q6	سهولت درک داده/ قابلیت درک داده	۰/۱۵۸	۰/۰۱۳	۰/۱۲۶	۰/۵۸۹	۰/۰۳۷	۰/۲۲۴
q14	امنیت/ انحصار دسترسی به داده	۰/۱۴۱	۰/۳۵۵	۰/۳۲۴	۰/۵۳۹	- ۰/۱۵۱	۰/۱۳۱
q15	ارزش افزوده داده	۰/۲۴۱	۰/۱۵۰	۰/۱۶۴	۰/۳۵۸	۰/۳۱۶	۰/۲۳۹
q16	قابلیت ارائه فشرده و مختصر	- ۰/۰۲۶	۰/۱۸۴	- ۰/۰۰۲	۰/۰۶۸	۰/۷۰۴	۰/۰۲۰
q12	مربوط بودن داده	۰/۴۳۸	۰/۰۸۵	۰/۱۳۹	۰/۱۳۰	۰/۴۶۵	۰/۱۰۴
q1	دقت و صائب بودن داده‌ها	۰/۰۳۳	- ۰/۰۲۱	۰/۳۳۵	۰/۰۱۶	۰/۰۳۰	۰/۶۲۷
q13	میزان سهولت بهره‌مندی از داده در دسترس	۰/۱۳۳	۰/۰۴۷	۰/۰۲۸	۰/۳۸۷	۰/۳۰۱	۰/۵۵۲

جدول ۳ داده‌های ماتریس دوران یافته تحت چرخش واریماکس مربوط به شناسایی و تأیید ویژگی‌های داده در کشف دانش را نشان می‌دهد. این داده‌ها نشان می‌دهد بالاترین بار چهار گویه اول بر روی عامل اول یعنی (q9، q18، q8، q10) است همچنین در دسته‌بندی سایر گویه‌ها در عامل‌های دوم تا ششم نیز، وجود این همبستگی بین گویه‌ها و عامل‌ها ملاحظه می‌شود و می‌توان بر اساس این همبستگی و بزرگترین بار عاملی، آن‌ها را در عامل‌ها دسته‌بندی کرد.

برای استخراج عنوان عامل در این بخش از پژوهش، مجدداً از رأی و نظر خبرگان استفاده شده و بر اساس مطالعات گروهی برای عوامل کشف شده در هر دسته از متغیرها عنوان مناسب انتخاب شده است. از آنجا که متغیرهای دارای همبستگی بیشتر با هریک از عامل‌ها در عامل مورد نظر قرار می‌گیرند، بنابراین متغیرهای مختلف در شش عامل مطابق آنچه که در جدول ۴ دیده می‌شود، نام‌گذاری شده است.

جدول ٤ . نام‌گذاری عوامل شناسایی و تأیید ویژگی‌های داده

متغیرها	نام عامل
معقولیت داده /قابل قبول بودن داده	سنجش کیفیت داده
دسترسی پذیری داده	
اعتبار داده	
مقبولیت منبع داده	
قابلیت تفسیر داده	سنجش پایداری داده
ابعاد مرتبط با زمان/ زمان محور بودن داده	
ثبات در سراسر منابع داده قابلیت اطمینان	
پاسخگویی داده/ زمان پاسخ داده	

کامل بودن و جامعیت داده‌ها	سنجش سازگاری داده
عینی و مشهود بودن داده	
قابلیت اتکا داده	
حجم مناسب داده/ حجم داده	
قابلیت انتقال داده	سنجش ارزشمندی و دسترسی‌پذیری داده
سهولت درک داده/ قابلیت درک داده	
امنیت/ انحصار دسترسی به داده	
ارزش افزوده داده	
قابلیت ارائه فشرده و مختصر	سنجش کارایی داده
مربوط بودن داده	
دقت و صائب بودن داده‌ها	سنجش صحت و کامل بودن
میزان سهولت بهره‌مندی از داده در دسترس	

پس از دسته‌بندی و مشخص شدن عوامل، به منظور اعتباریابی مدل پژوهش از تحلیل عاملی تأییدی و الگوسازی معادلات ساختاری با استفاده از نرم‌افزار Lisrel 8.80 استفاده شده است. شکل ۱، خروجی تحلیل عاملی تأییدی در حالت تخمین استاندارد را نشان می‌دهد. هر چه بار عاملی به عدد یک نزدیکتر باشد، یعنی مؤلفه بهتر می‌تواند متغیر مکنون را تبیین نماید. اگر بار عاملی کمتر از ۰/۳ باشد رابطه ضعیف در نظر گرفته می‌شود، و از آن صرفنظر می‌شود، بار عاملی بین ۰/۳ تا ۰/۶ قابل قبول و اگر بیش از ۰/۶ باشد کاملاً مطلوب است. همچنین مقدار شاخص RMSEA [1] کمتر از ۰/۰۵ نشان دهنده برازش خوب مدل است. همانطور که در شکل ۱ و

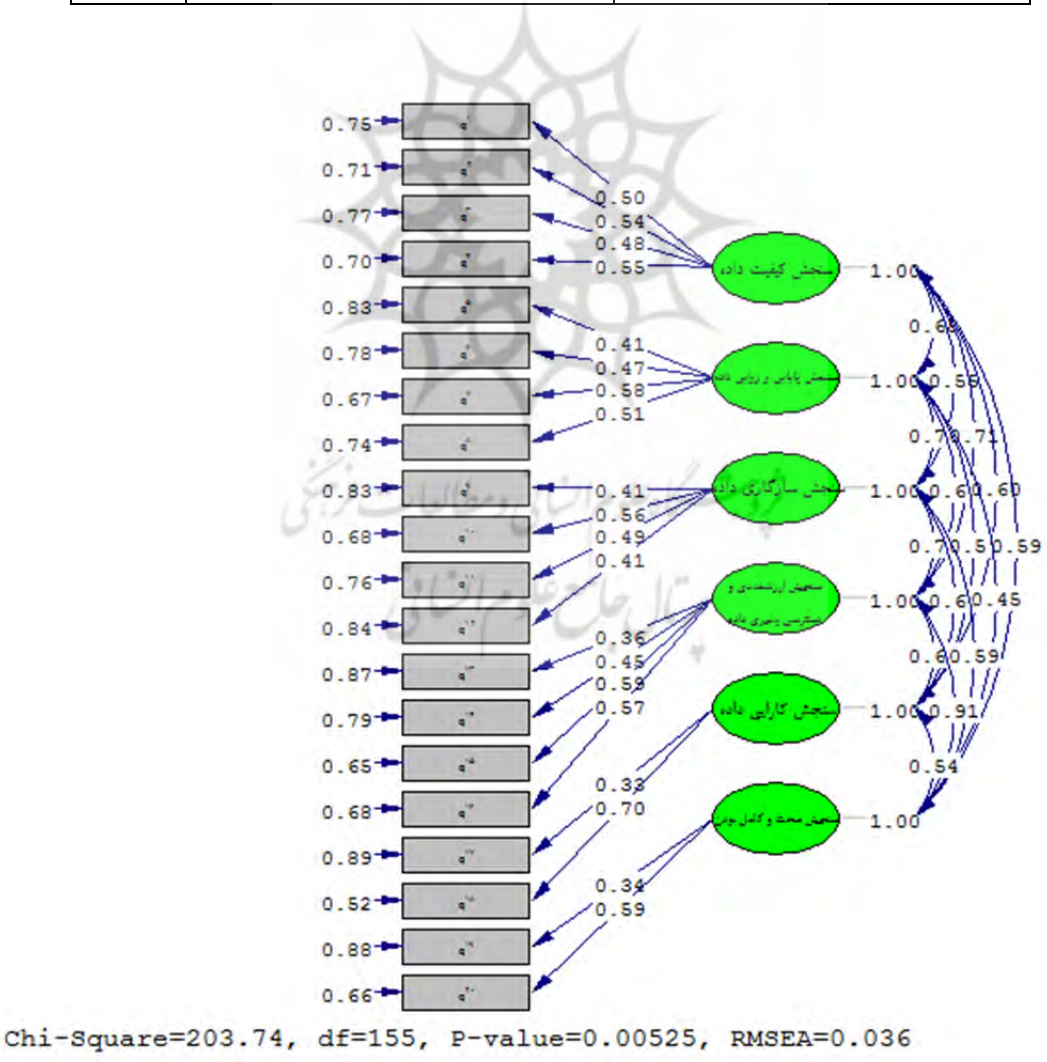
جدول ۵ دیده می‌شود، بار عاملی همه مولفه‌ها بیشتر از ۰/۳ است. بنابراین، این مؤلفه‌ها تبیین کننده مناسبی برای متغیر مورد نظر هستند و شاخص RMSEA هم برابر ۰/۰۳۶ است که شرط مورد نظر برازش را ارضا می‌نماید.

جدول ۵. ضرایب بارهای عاملی مؤلفه‌های پژوهش

مؤلفه‌ها	شاخص	بار عاملی
سنجش کیفیت داده	معقولیت داده /قابل قبول بودن داده	۰/۵۰
	دسترسی‌پذیری داده	۰/۵۴
	اعتبار داده	۰/۴۸
	مقبولیت منبع داده	۰/۵۵
سنجش پایایی و روایی داده	قابلیت تفسیر داده	۰/۴۱
	ابعاد مرتبط با زمان/ زمان محور بودن داده	۰/۴۷
	ثبات در سراسر منابع داده قابلیت اطمینان	۰/۵۸

1 Root mean square error of approximation) Weighted root-mean-square residual

	پاسخگویی داده/ زمان پاسخ داده	۰/۵۱
سنجش سازگاری داده	کامل بودن و جامعیت داده‌ها	۰/۴۱
	عینی و مشهود بودن داده	۰/۵۶
	قابلیت اتکا داده	۰/۴۹
	حجم مناسب داده/ حجم داده	۰/۴۱
سنجش ارزشمندی و دسترسی‌پذیری داده	قابلیت انتقال داده	۰/۳۶
	سهولت درک داده/ قابلیت درک داده	۰/۴۵
	امنیت/ انحصار دسترسی به داده	۰/۵۹
	ارزش افزوده داده	۰/۵۷
سنجش کارایی داده	قابلیت ارائه فشرده و مختصر	۰/۳۳
	مربوط بودن داده	۰/۷
سنجش صحت و کامل بودن	دقت و صائب بودن داده‌ها	۰/۳۴
	میزان سهولت بهره‌مندی از داده در دسترس	۰/۵۹



شکل ۱. مدل اندازه‌گیری عوامل مؤثر رفتاری بر ارتباطات علمی در حالت تخمین استاندارد

یافته‌ها

پس از تکمیل نتایج محاسبات و آزمون‌های آماری می‌توان یافته‌های پژوهش را گزارش نمود. بر اساس داده‌های گردآوری شده، مشخصات دموگرافیک پاسخ‌دهندگان به پرسشنامه در جدول ۶ آمده است.

جدول ٦. مشخصات دموگرافیک پاسخ‌دهندگان

مشخصه	تعداد		درصد
سن	۲۰ تا ۳۰ سال	۷۳	۲۹/۷
	۳۰ تا ۴۰ سال	۱۴۰	۵۶/۸
	۴۰ تا ۵۰ سال	۳۳	۱۳/۵
	۵۰ تا ۶۰ سال	۱	کمتر از یک درصد
	بالای ۶۰ سال	۰	۰
رشته تخصصی یا شغلی	متخصص داده کاوی و علوم داده	۱۳۹	۵۶/۲
	متخصص یادگیری ماشین	۴۰	۱۶/۱
	متخصص تحلیل کلان داده	۳۷	۱۴/۹
	متخصص فناوری اطلاعات سلامت	۲۵	۱۰/۱
	همه گزینه‌ها	۶	۲/۴
مدت فعالیت در حوزه تخصصی	کمتر از ۵ سال	۱۳۳	۵۴/۱
	۵ تا ۱۰ سال	۷۳	۲۹/۷
	۱۰ تا ۲۰ سال	۳۴	۱۳/۵
	بیش از ۲۰ سال	۷	۲/۷

در پاسخ به سوال‌های پژوهش و بر اساس نتایج آزمون تحلیل عاملی اکتشافی و تأییدی در رابطه با اثرگذاری عامل " کیفیت داده"، می‌توان گفت نتایج آزمون‌ها نشان‌دهنده آن است که این عامل با چهار سؤال در پرسشنامه، و بار عاملی بالای ۰/۳ برای همه مؤلفه‌های این عامل، به درستی شناخته شده و این عامل از ویژگی‌های مورد نظر و مطلوب داده جهت کشف دانش در سیستم‌های کلان داده سلامت است. همچنین بار عاملی عامل "دسترسی پذیری داده" بر کشف دانش از سیستم‌های کلان داده سلامت، و عامل کامل بودن داده‌ها، همچنین مولفه‌های عامل کارایی داده (دو مؤلفه) و همه چهار مؤلفه عامل پایداری داده، فاکتورهای مربوط به عامل سازگاری داده و مؤلفه‌های پایداری داده بالاتر از ۰/۳ بوده و بنابراین عوامل شناسایی شده همگی از ویژگی‌های مورد نظر و مطلوب داده جهت کشف دانش در سیستم‌های کلان داده سلامت است.

همچنین جهت پاسخ به سؤال دیگر پژوهش می‌توان گفت بر مبنای شاخص‌های برازش مدل بر اساس p-valueمحاسبه شده توسط نرم‌افزار که کمتر از ۰/۰۵ است و شاخص RMSEA که برابر ۰/۳۶ است و میزان مطلوب آن کمتر از ۰/۰۸ است، مدل اکتشافی مورد تأیید قرار گرفته است و بنابراین عامل‌های دقت و کیفیت داده‌ها، دسترسی‌پذیری داده، کامل بودن داده‌ها، کارایی داده‌ها، پایداری داده‌ها و سازگاری داده ویژگی‌های داده مطلوب برای استفاده در فرایند کشف دانش از سیستم‌های کلان داده است.

بحث و نتیجه گیری

در راستای دستیابی به هدف غایی و همیشگی بقاء، انسان همواره در پی راه‌کارهایی برای بهبود وضعیت سلامت و یافتن شیوه‌های درمانی مؤثر برای بیماری‌ها بوده است. ظهور و بروز ابزارهای فناوری اطلاعات از پرونده الکترونیک سلامت گرفته تا سیستم‌های کلان داده و کشف دانش از آنها نیز به حرکت در مسیر دستیابی به این هدف کمک شایانی نموده است. استخراج نظام‌مند دانش مفید از مجموعه داده‌های عظیم که در منابع مختلف و متفاوت ذخیره شده‌اند و کشف الگوهای پنهان در آنها، جهت درک بهتر محیط پیرامون و اخذ تصمیم بهینه در زمان معقول، از اهمیت ویژه‌ای برخوردار است. در حوزه مراقبت‌های بهداشتی و سلامت، پزشکان، ارائه‌دهندگان و تأمین‌کنندگان مراقبت‌های بهداشتی، سیاست‌گذاران و مهمتر از همه بیماران فرصت‌های ارزشمندی را در پرتو دانش و اطلاعات استخراج شده از کلان داده بدست آورده‌اند. از جمله این فرصت‌ها می‌توان به تشخیص، درمان و مراقبت پزشکی و حمایتی بهنگام، کاهش هزینه‌های درمان و مراقبت‌های بهداشتی، توسعه پزشکی دقیق، کنترل و مقابله با بیماری‌های فراگیر همچون کووید ۱۹، سیاست‌گذاری مبتنی بر واقعیت، ارائه خدمات درمانی جهان شمول، مقابله با بیماری‌های غیرواگیر، و موارد پرشمار دیگری اشاره کرد. با این حال دستیابی به مزایای کشف دانش از مجموعه‌های متنوع، حساس، پیچیده، ساختارمند و فاقد ساختار کلان داده حوزه سلامت با چالش‌های قابل توجهی روبرو است. از جمله این چالش‌ها فراهم‌سازی داده‌های درخور برای سیستم‌های پردازش کلان داده سلامت است. فراهم‌سازی کلان داده‌های با کیفیت و در خور، فرایند انتخاب الگوریتم‌های پردازش و سیستم‌های پردازش مناسب را ممکن می‌سازد. در این راستا، پژوهش حاضر با بهره‌گیری از مرور ادبیات و مصاحبه عمیق با خبرگان اقدام به پاسخگویی دو سؤال اصلی کرد. الف) ویژگی‌های داده درخور در حوزه سلامت جهت کشف بهینه دانش از آنها کدامند؟ ب) آیا ویژگی‌های داده درخور بر فرایند کشف دانش از سیستم‌های کلان داده سلامت تأثیر دارد؟ پژوهشگر با بهره‌گیری از داده‌های جمع‌آوری شده جهت پاسخگویی به سؤال اول پژوهش از روش تحلیل عاملی استفاده نمود که در نهایت به شش ویژگی اصلی داده درخور شناسایی و تعیین گردید. این ویژگی‌ها عبارتند از: دقت و کیفیت داده، دسترسی‌پذیری داده، کامل بودن داده، کارایی داده، پایداری داده و سازگاری داده. نتایج این بخش از پژوهش همراستا با ویژگی‌هایی است که سازمان بهداشت جهانی (۲۰۰۳) و مؤسسه اطلاعات سلامت کانادا (۲۰۱۱) به عنوان ویژگی‌های داده مطلوب برای استفاده از سیستم‌های اطلاعات سلامت منتشر کرده‌اند. همچنین بنابر نظر خبرگان حوزه کلان داده سلامت، ویژگی‌های ذکر شده فرایند کشف دانش از آنها در سیستم‌های پردازش کلان داده سلامت را بهبود می‌بخشد. ویژگی‌های شناسایی شده ضمن قابلیت تعمیم برای کلیه مجموعه کلان داده در سیستم‌های پردازشی مختلف، با تأکید بر عدم وابستگی به فرمت، دارای خصایصی فراگیر است که همه حوزه‌ها از جمله حوزه کلان داده سلامت را دربر می‌گیرد. با توجه به تأیید اعتبار و برازش مطلوب مدل پیشنهادی پژوهش حاضر جهت شناسایی کلان داده در خور حوزه سلامت، این مدل می‌تواند ابزار مناسبی برای سنجش ویژگی‌های داده مناسب جهت کشف دانش از سیستم‌های کلان داده سلامت باشد. همچنین نتایج پژوهش حاضر از نظر اعتبار همتایان توسط پژوهش‌های انجام شده توسط وانگ (۲۰۱۵) و دان وو (۲۰۲۱) مورد تأیید قرار می‌گیرد.

در تبیین و نتیجه‌گیری کلی از نتایج پژوهش حاضر می‌توان به این موضوع مهم اشاره نمود که دقت و حساسیت در به‌کارگیری داده-های درخور و مناسب جهت کشف دانش از کلان داده سیستم‌های سلامت متضمن کارایی و اثربخشی دانش استخراج شده از این سیستم‌ها و در نتیجه بهبود و ارتقاء سلامت، بهداشت و کیفیت زندگی در سطح فردی و اجتماعی است. همچنین این بحث ظریف می‌تواند راه‌گشای تصمیمات و سیاست‌های دقیق و صحیح در حوزه سلامت، بهداشت و درمان شود و در مواقع بحرانی همچون تجربه اخیر شیوع همه‌گیر بیماری کرونا به یاری جوامع و سیاست‌گذاران برخیزد. در جهت توسعه دانش این حوزه و با توجه یافته‌های حاصل از فرایند و نتایج و دامنه پژوهش حاضر پیشنهاد می‌شود پژوهشگران پژوهش‌های آتی خود را بر ابعاد امنیت و محرمانگی کلان داده حوزه سلامت، مهندسی کلان داده سلامت، ارتباط کلان داده سلامت و علم داده، و برخورد با اختلال و عدم قطعیت در کلان داده سلامت معطوف نمایند. همچنین پیشنهاد می‌شود ابعاد مورد نیاز جهت تعیین اولویت‌های پاکسازی و پیش پردازش داده‌های سیستم‌های کلان داده سلامت جهت کشف دانش از آنها و نهایتاً ارائه مسیری برای انجام این مهم توسط پژوهشگران بررسی گردد.

منابع

احمدی، م.، دهقانی محمودآبادی، آ.، و فزونخواه، ش. (۱۳۹۰). استاندارد تبادل پیام پرونده‌ی الکترونیک سلامت در سازمان‌های منتخب. *مدیریت اطلاعات سلامت*، ۹(۲)، ۱۶۱-۱۷۱.

کیخا ل.، صفدری ر.، قاضی سعیدیم.، سیدفرج اله س. ص.، و محمدزاده ن. (۱۳۹۵). استانداردها، زیرساخت توسعه پرونده الکترونیک سلامت ملی: مطالعه تطبیقی. *مجله انفورماتیک سلامت و زیست پزشکی*، ۳ (۳)، ۲۲۲-۲۱۴

مقدسی، ح. (۱۳۸۴). *کیفیت اطلاعات در مراقبت بهداشتی*. انتشارات واژه پرداز.

مقدسی، ح.، و قائمی، م. (۱۳۹۴). مطالعه تطبیقی سه استاندارد امنیت داده در نظام سلامت. *مجله انفورماتیک سلامت و زیست پزشکی*، ۲ (۳)، ۱۸۴-۱۹۴

مقدسی، ح.، و عیانی، ش. (۱۳۹۲). امنیت داده‌ها در سیستم‌های اطلاعات سلامت. *فصلنامه پژوهش‌های حفاظتی و امنیتی*، ۲(۶).

میرحسینی، ز.، رضایی، ز.، و سپهر، ف. (۱۴۰۰). کاربرد کلان داده در کتابخانه‌های عمومی: تعیین ضرورت‌ها و عوامل تاثیرگذار. *دانش شناسی (علوم کتابداری و اطلاع رسانی و فناوری اطلاعات)*، ۹۶-۱۱۵.

نظری، ج.، و مختاری پور، م. (۱۳۸۸). تحلیل عاملی و کاربرد آن در علوم اجتماعی. *کتاب ماه علوم اجتماعی*، ۲۰-۳۳.

References

Abbasi, A., Sarker, S., & Chiang, R. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, 17(2), i – xxxii.

Abdullah, N., Ismail, S., Sophiayati, S., & Mohd Sam, S. (2015). Data quality in big data: a review. *International journal Advance Soft Computer*, 7(3), 17-27.

Ahmadi, M., Dehghani Mahmoodabadi, A., & Foozonkhah, Sh. (2012). Messaging Standards for Electronic Health Records in Selected Organizations. *Health Information Management*, 9(2), 171 [In Persian]

Alfred, R. (2005). Knowledge Discovery: Enhancing Data Mining and Decision Support Integration. *The University of York*.

Alsunaidi , S., Almuhaideb, A., Ibrahim, N., Shaikh, F., Alqudaihi, K., Alhaidari , F., & Alahrani, M. (2021). Applications of Big Data Analytics to Control COVID-19 Pandemic. *Sensors*, 21(7), 2282.

Brown, N. (2015, Sep 11). Healthcare Data Growth: An Exponential Problem. (Nextech) Retrieved from Jan 07, 2022 ، ازhttps://www.nextech.com/blog/healthcare-data-growth-an-exponential-problem

Cios, K. J., & Kacprzyk, J. (2001). Medical data mining and knowledge discovery. *Physica-Verlag*.

Cios, K., & Moore, G. (2002). Uniqueness of medical data mining. *Artificial Intelligence in Medicine*, 26(1-2), 1-24.

Galetsi, Panagiota, Korina Katsaliaki, & Sameer Kumar (2020). Big Data Analytics in Health Sector: Theoretical Framework, Techniques and Prospects. *International Journal of Information Management*, 50.

Hong, Liang et al. (2018). Big Data in Health Care: Applications and Challenges. *Data and Information Management*, 2(3).

Jee, Kyoungyoung, & Gang Hoon K. (2013). Potentiality of Big Data in the Medical Sector: Focus on How to Reshape the Healthcare System. *Healthcare Informatics Research*, 19(2).

Keikha L, Safdari R, Ghazisaeedi M, Seiedfarajollah SS, Mohammadzadeh N. (2016). Standards as Infrastructure for Nationa Electronic Health Record Development: A Comparative Study. *Journal of Health and Biomedical Informatics*, 3(3), 214-222. [In Persian]

Lawton, G. (2021, Apr 27). Data quality for big data: Why it's a must and how to improve it. (TechTarget) Retrieved from Jan 14, 2022 ،https://searchdatamanagement.techtarget.com/feature/Data-quality-for-big-data-Why-its-a-must-and-how-to-improve-it

Li, Y., Thomas, M., & Osei-Bryson, K.-M. (2016). A snail shell process model for knowledge discovery via data analytics. *Decision Support Systems*, 91, 1-12.

Mirhosseini, Z., Rezaie, Z., Sepehr, F. (2021). Application of Big data in the Public Libraries: Determining the Necessities and Affecting Factors. *Journal of Knowledge Studies*, 14(53), 96-115. [In Persian]

Moghadesi, H. (2020). Quality of information in health care. Tehran: *Vazheh Pardaz*. [In Persian]

Moghadesi, H., & Ghaeme, MM. (2015). A Comparative Study of Three Standards of Data Security in Health Systems. *Journal of Health and Biomedical Informatics*, 2(3), 184-194. [In Persian]

Moghadesi, H., & Ayani, S. (2013). Data security in health information systems. Quarterly *Journal of Security Protection Research, Imam Hussein University (PBUH)*, 127-137. [In Persian]

Nazari, J., & Mokhtari, M. (2009). Factor analysis and its application in social sciences. *Social Science Book of Month*, 20-33. [In Persian]

O'Driscoll, A., Daugelaite, J., & Sleator, R. (2013). Big data',Hadoop and cloud computing in genomics. *Journal of biomedical informatics*, 46(5), 774-781.

Pramanik, P. K., Pal, S., & Mukhopadhyay, M. (2022). Healthcare big data: A comprehensive overview. *Research Anthology on Big Data Analytics, Architectures, and Applications* (P. 119-147).

Prasser, F., Kohlmayer, F., Spengler, H., & Kuhn, K. (20117). A scalable and pragmatic method for the safe sharing of high-quality health data. *IEEE journal of biomedical and health informatics*, 2, 611-622.

Pribor, H. (1989). Expert systems in laboratory medicine: A practical consultative application. *Journal of Medical Systems*, 13(2), 103-109.

Roddick, J., Fule, P., & Graco, W. (2003). Exploratory medical knowledge discovery: Experiences and issues. *ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Explorations*, 5(1), 94-99.

Shobana, V., & N. Kumar. (2015). Big Data - A Review. *International Journal of Applied Engineering Research*, 10(55).

Wong, H. T., Yin, Q., Guo, Y., Murray, K., Zhou, D., & Slade, D. (2015). Big data as a new approach in emergency medicine research. *Journal of Acute Disease*, 4(3), 178-179.

Wu, D., Xu, H., Yongyi, W., & Zhu, H. (2021). Quality of government health data in COVID-19: definition and testing of an open government health data quality evaluation framework. *Library Hi Tech*.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی