Journal of Population and Development

3

Vol. 2
Spring 2024

Research Paper

Received:
08 May 2024
Revised:
20 May 2024
Accepted:
21 July 2024
Published:
21 July 2024
P.P: 71-100

Population Data Science Dashboard; What, Why and How

Yaser Khorrami [1]

Abstract

The use of data covering human populations is increasingly used for research and decision-making. However, dealing with these data in a scientific manner, which has the specific characteristics of the population in contrast to other types of data, is a subject that has been neglected in the age of data science and big data. Population data science can be designed as an interdisciplinary form of combining data science and population science with a specialized view of population data. Since the full potential of population data can often only be achieved when such data are linked to other databases, a comprehensive and basic look at the subject of population data science is one of the basic requirements in the field of population. In this article, it has tried to show what and why this science is, how to implement the conceptual architecture based on big data and cloud computing, relying on the current knowledge of the field of data science in the form of a proposed management dashboard model of population data science.

Keywords: Population data, Cloud computing, Population data science, Data science, Big data, Management dashboard model

1. Corresponding Author: PhD student of regional studies, Imam Hossein University, Tehran, Iran.
miladpahlavan75@gmail.com

Cite this Paper: Khorrami' Y. Population Data Science Dashboard; What, Why and How, 3(2), 71–100.

Publisher: Imam Hussein University

© **Authors**

This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0) .

نشریه جمعیت و پیشرفت

۳

سال دوم
بهار ۱۴۰۳

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۰۲/۱۹
تاریخ بازنگری: ۱۴۰۳/۰۲/۳۱
تاریخ پذیرش: ۱۴۰۳/۰۴/۳۱
تاریخ انتشار: ۱۴۰۳/۰۶/۱۳
صص: ۱۰۰-۷۱

داشبورد علم داده جمعیتی؛ چیستی، چرایی و چگونگی

یاسر خرّمی [۱]

چکیده

استفاده از داده‌هایی که جمعیت‌های انسانی را پوشش می‌دهند به طور فزاینده‌ای برای تحقیق و تصمیم‌گیری در حال افزایش است. با این حال، پرداختن به این داده‌ها بصورت علمی، که مشخصه اختصاصی جمعیت را در تفاوت با سایر انواع داده‌ها دارند، موضوعی است که در عصر علم داده و کلان‌داده، بصورت تخصصی کماکان مغفول مانده است. علم داده‌های جمعیتی بصورت میان رشته‌ای از تجمیع علم داده و علم جمعیت با نگاه تخصصی به داده‌های جمعیتی قابل طرح است. از آنجاییکه پتانسیل کامل داده‌های جمعیت اغلب تنها زمانی قابل دستیابی است که چنین داده‌هایی به پایگاه‌های داده دیگر مرتبط شوند، نگاه کلی‌نگر و اساسی به موضوع علم داده‌های جمعیتی یکی از ملزومات اساسی در حوزه جمعیت است. در این مقاله تلاش شده است تا با بیان چیستی و چرایی این علم، چگونگی پیاده‌سازی معماری مفهومی مبتنی بر کلان‌داده و رایانش ابری با تکیه بر دانش روز حوزه علم داده در قالب مدل پیشنهادی داشبورد مدیریتی علم داده جمعیتی نشان داده شود.

کلیدواژه‌ها: داده‌های جمعیتی، رایانش ابری، علم داده‌های جمعیتی، علم داده، کلان‌داده، مدل داشبورد مدیریتی

۱. استادیار، دانشکده پیامبر(ص)، دانشگاه جامع امام حسین(ع)، تهران، ایران
y.khorrami@ihu.ac.ir

استناد: خرّمی، یاسر. داشبورد علم داده جمعیتی؛ چیستی، چرایی و چگونگی ، ۳(۲)، ۱۰۰-۷۱.

ناشر: دانشگاه جامع امام حسین (ع)

© نویسندگان

این مقاله تحت لیسانس آفرینندگی مردمی (Creative Commons License- CC BY) در دسترس شما قرار گرفته است.

۱- مقدمه

جهان به سرعت در حال گذار از حالت سنتی به دیجیتالی است و این موضوع خود را در تبدیل فزاینده داده‌ها به حالت دیجیتالی نشان می‌دهد (Berntzen et al., 2018). در عین حال انباشتگی داده فرصت‌هایی را به منظور شناسایی و حل مسائل مربوطه به حوزه‌های مختلف بشری ایجاد می‌کند و حتی فراتر از این موضوع، باعث شکل‌گیری نوآوری در عرصه‌های مختلف دانشی خواهد شد (Ashford et al., 2022). یکی از این حوزه‌ها، پاسخگویی به مسائل جامعه انسانی از طریق تمرکز بر موضوع جمعیت‌های انسانی است. با استفاده از مجموعه داده‌های مرتبط با موضوع جمعیت که به اندازه کافی بزرگ، پیچیده و متنوع باشند، می‌توان در زمان مناسب نسبت به ایجاد دانش مرتبط با موضوع جمعیت اقدام نمود. علم داده‌های جمعیتی[1] یک حوزه چندرشته‌ای است که شامل جمع‌آوری، تجزیه و تحلیل، تفسیر و ارائه داده‌های مربوط به جمعیت‌های انسانی می‌شود. این علم به‌عنوان علم داده‌های مربوط به افراد[2] بصورت ترکیبی از دو حوزه علم‌داده و علم جمعیت[3] قابل تعریف است (McGrail et al., 2018).

علم داده‌های جمعیتی یک رشته نسبتاً جدید است که هنوز در مراحل اولیه رشد و بالندگی است. دلایل متعددی وجود دارد که چرا این علم هنوز به طور کامل توسعه نیافته است، از جمله این دلایل عبارت‌اند از:

- **در دسترس بودن داده‌ها**: دستیابی به داده‌های جمعیت به ویژه برای کشورهای در حال توسعه اغلب دشوار است. این امر می‌تواند انجام مطالعات در مقیاس بزرگ را برای محققان دشوار کند.
- **کیفیت داده‌ها**: کیفیت داده‌های جمعیت می‌تواند به طور قابل توجهی متفاوت از هم باشد. این امر باعث دشواریِ رسیدن به نتیجه‌گیری دقیق از مطالعات برای محققان خواهد شد.
- **پیچیدگی داده‌ها**: داده‌های جمعیت اغلب پیچیده و تجزیه و تحلیل آنها دشوار است. این امر می‌تواند شناسایی الگوها و روندهای معنادار را برای محققان دشوار نماید.

1. Population Data Science
2. The science of data about people
3. Population science

- **تخصص ناکافی**: کمبود تخصص در علم داده‌های جمعیتی وجود دارد. این امر می‌تواند انجام مطالعات مناسب از منظر روش‌شناختی و دستیابی به نتایج قابل اعتمادی را دشوار نماید.
- **تامین مالی**: علم داده‌های جمعیتی رشته‌ای با بودجه مناسب نیست. بنابراین ممکن است انجام مطالعات در مقیاس بزرگ و طولانی مدت را برای محققان دشوار کند.

علیرغم این چالش‌ها، علم داده‌های جمعیتی یک حوزه رو به رشد است. همانطور که داده‌های بیشتری در دسترس قرار می‌گیرد و محققان تجربه بیشتری در تجزیه و تحلیل آن به دست می‌آورند، می‌توان انتظار داشت که شاهد پیشرفت بیشتری در این زمینه باشیم. علاوه بر موارد فوق باید اذعان نمود که با توجه به اینکه علم داده‌های جمعیتی یک زمینه چند رشته‌ای است، تعداد کمی از افراد دارای تمام دانش و مهارت‌های مورد نیاز علم داده‌های جمعیت هستند. پیامدهای امکان‌سنجی و ارائه این علم بصورت یک مجموعه دانشی به هم پیوسته، مستلزم پیشرفت جنبه‌های مختلف آن، نظیر فناوری داده، اشکال مختلف تجزیه و تحلیل داده، و روش‌های مشارکت علم جمعیت‌شناسی و ... است که نقش مهمی در درک روندها و الگوهای جمعیتی مختلف ایفا می‌کند و اثرات زیادی برای جهت‌گیری در سیاست‌های عمومی، برنامه‌ریزی اجتماعی و توسعه اقتصادی دارد. این پیامدها سرآغاز یک دستور کار تحقیقاتی برای علم داده‌های جمعیت است که اگر به عنوان یک زمینه جمعی مورد بررسی قرار گیرد، می‌تواند پیشرفت‌های قابل توجهی را در درک ما از جمعیت نظیر روندهای رشد موالید، جامعه، سلامت و رفتار انسانی ایجاد کند.

خانم دکتر بوهو از دانشگاه تنسی بیان می‌کند: "در حال حاضر، آموزش روش‌شناسی، آمار و رایانه ما به عنوان جمعیت‌شناس، ما را برای مقابله با انواع مشکلاتی که می‌توان با این نوع داده‌های بزرگ برطرف کرد، آماده نکرده است. یعنی ما داده‌های بزرگ جمعیت را داریم، اما به شکل کوچکی با آن رفتار می‌کنیم. حتی اگر می‌دانستیم چگونه چهار دهه داده‌های سرشماری ایالات متحده را در سیستمی با حافظه کافی یا پردازنده‌های کافی جمع‌آوری کنیم، آیا می‌دانیم با آن چه کنیم؟ بسیاری از ما این کار را نمی‌کنیم. بنابراین تجزیه داده‌ها به قطعات ساده اما کوچک قابل استفاده است، اما انجام این کار از تجسم کامل داده‌ها نیز جلوگیری می‌کند. چه چیزی را از دست داده‌ایم؟ در اینجا، من از مرز جدید جمعیت‌شناسی بحث می‌کنم زیرا شامل پیشرفت در محاسبات و در دسترس بودن تکنیک‌های جدید برای ترکیب داده‌ها است" (Bohon, 2018).

این مقاله تلاش دارد تا از موضـوع نسـبتاً بدیعی تحت عنوان علم داده‌های جمعیتی صـحبت نموده و چیسـتی، چرایی و چگونگی آن را به بحث بگذارد. بنابراین پس از بیان تفاوت‌های علم جمعیت و جمعیت‌شـناسـی در بخش دوم، به بیان چیسـتی علم داده‌های جمعیتی در بخش سـوم پرداخته و متعاقب آن در بخش چهارم به چرایی این علم توجه خواهد شـد. بخش پنجم پیشـینه موضـوع و نمونه‌های موفق پیاده‌شـده از آن را در جهان شـرح می‌دهد. چگونگی تحقق این علم با بیان مدل معماری مفهومی مبتنی بر کلان‌داده و یادگیری ماشین/هوش مصنوعی در فضای رایانش ابری با هدف تحقق داشـبورد مدیریتی علم داده‌های جمعیت و ارائه مهمترین فناوری‌های موجود برای پیاده‌سازی آن در فصل ششم بیان خواهد شد.

۲- علم جمعیت و جمعیت‌شناسی

جمعیت[1] بیانگر تعداد کل افراد در یک منطقه جغرافیایی معین در یک زمان معین اسـت در حالیکه جمعیت‌شناسی[2]، دربرگیرنده مطالعه ویژگی‌های جمعیت‌های انسانی است. علم جمعیت و علم جمعیت‌شناسی، رشته‌های مرتبطی هستند که هر دو بر مطالعه جمعیت‌های انسانی و ویژگی‌های آنها تمرکز دارند، اما تأکیدات و دامنه‌های آنها کمی با هم متفاوت است.

علم جمعیت[3] یک رشته چندوجهی است که به مطالعه جامع جمعیت‌های انسانی و ویژگی‌های پیچیده آنها اختصاص دارد. این ویژگی‌ها نه تنها جنبه‌های اساسیِ اندازه، رشد، توزیع و ترکیب را در بر می‌گیرند، بلکه به عوامل بی‌شـمار مؤثر در این پویایی‌ها، از جمله باروری، مرگ‌ومیر و مهاجرت نیز می‌پردازند. به عنوان یک رشـته واقعاً بین رشـته‌ای، علم جمعیت از طیف متنوعی از رشته‌ها، از جمله جمعیت‌شناسی، اقتصاد، جامعه‌شناسی و بهداشت عمومی، اما نه محدود به آن، استفاده می‌کند.

جمعیت‌شـناسـی در لغت‌نامه آکسـفورد به صـورت «داده‌های مربوط به جمعیت و گروه‌های مختلف درون آن»[4] تعریف شـده اسـت. علم جمعیت‌شـناسـی[5] که اغلب به سـادگی به عنوان

1. Population
2. Demographics
3. Population science
4. Data relating to the population and different groups within it
5. Demography science

جمعیت‌شناسی شناخته می‌شود، مطالعه ویژگی‌های آماری جمعیت‌ها، مانند نرخ تولد، نرخ مرگ و میر، الگوهای مهاجرت و عوامل مختلف دیگری است که بر اندازه، ترکیب و توزیع جمعیت‌ها تأثیر می‌گذارد. جمعیت‌شناسان این داده‌ها را برای درک روندها و الگوهای پویایی جمعیت تجزیه و تحلیل می‌کنند. جمعیت‌شناسی در درجه اول با جنبه های کمّی جمعیت از جمله رشد، ساختار و حرکت آنها سروکار دارد. همچنین شامل توسعه مدل‌ها و نظریه‌های جمعیت‌شناختی برای توضیح تغییرات جمعیت است.

علم جمعیت، یک حوزه میان رشته‌ای گسترده‌تر است که فراتر از تجزیه و تحلیل آماری داده‌های جمعیتی را در بر می‌گیرد و مفاهیم و روش‌های رشته‌های مختلف از جمله جمعیت‌شناسی، جامعه‌شناسی، اپیدمیولوژی، مردم‌شناسی، بهداشت عمومی، اقتصاد و غیره را با هم ادغام می‌کند. هدف علم جمعیت شناخت نه تنها جنبه‌های کمّی جمعیت بلکه عوامل اجتماعی، اقتصادی، فرهنگی و مرتبط با سلامت است که پویایی جمعیت را هدایت و تبیین می‌کنند. این علم اغلب روابط و تعاملات پیچیده بین ویژگی‌های جمعیت و مسائل اجتماعی گسترده‌تر، مانند نابرابری‌های بهداشتی، اثرات زیست‌محیطی، و پیامدهای سیاست را نیز بررسی می‌کند.

در اصل، علم جمعیت‌شناسی، زیرمجموعه‌ای از علم جمعیت است که عمدتاً بر تجزیه و تحلیل آماری و کمّی داده‌های جمعیت تمرکز دارد. از سوی دیگر، علم جمعیت رویکردی کل نگرتر دارد و جنبه‌های چند بعدی جمعیت‌ها و تعامل آنها با عوامل مختلف اجتماعی را بررسی می‌کند. هر دو حوزه، بینش‌های ارزشمندی را برای درک چگونگی تغییر، تکامل و تأثیر جمعیت‌ها بر جهان اطراف خود ارائه می‌نمایند. دانشمندان جمعیت با بسیاری از موضوعات کلیدی شامل: اندازه و رشد جمعیت، توزیع جمعیت، ترکیب جمعیت و پویایی جمعیت درگیر می‌شوند تا پیچیدگی‌های جمعیت‌های انسانی را کشف نمایند.

برخی از موضوعات مورد بررسی در علم جمعیت‌شناسی عبارت‌اند از:

- نرخ خام موالید[1]: که شامل تعداد تولدهای زنده در یکسال معین به ازای هر ۱۰۰۰ نفر است.
- نرخ خام مرگ و میر[2]: تعداد مرگ و میر در یک‌سال معین به ازای هر ۱۰۰۰ نفر است.

1. Birth Rate
2. Death Rate (Mortality Rate)

- نرخ باروری[1]: میانگین تعداد فرزندانی که یک زن در طول زندگی خود به دنیا می‌آورد که براساس نرخ‌های زاد و ولد در سن ویژه باروری محاسبه می‌شود.
- امید به زندگی[2]: میانگین سال‌هایی که یک نوزاد متولد شده می‌تواند انتظار زندگی یا زنده بودن داشته باشد. در این تعریف فرض بر این است که میزان مرگ و میر فعلی در طول زندگی ثابت بماند.
- هرم جمعیتی[3]: یک نمایش گرافیکی از توزیع سنی و جنسیتی یک جمعیت است که به تجسم تعادل بین گروه‌های سنی مختلف کمک می‌کند و بینش‌هایی را در مورد پتانسیل رشد جمعیت، ساختار اجتماعی و چالش‌های آینده ارائه می‌دهد.
- مهاجرت: جابه‌جایی افراد از یک مکان به مکان دیگر، که در دو حالت مهاجرت داخلی (جابجایی در داخل مرزهای یک کشور) و مهاجرت بین‌المللی (جابه‌جایی بین کشورها)، قابل تعریف است و الگوی آن می‌تواند تاثیرات قابل‌توجهی را بر اندازه و ترکیب جمعیت و همچنین بر پویایی فرهنگی و اقتصادی داشته باشد (Subbotin and Aref, 2021).
- رشد و کاهش جمعیت[4]: رشد جمعیت براساس نرخ زاد و ولد، نرخ مرگ و میر و مهاجرت خالص (تفاوت بین مهاجرت داخلی و بین‌المللی) قابل اندازه‌گیری است. افزایش جمعیت منجر به افزایش تقاضای منابع و چالش‌های بالقوه در ارائه خدمات می‌شود در حالیکه کاهش جمعیت می‌تواند منجر به کمبود نیروی کار، پیری جمعیت، افزایش هزینه‌های سالمندی نظیر بیمه، کاهش توان نظامی کشورها و ... شود.
- نظریه گذار جمعیتی[5]: الگویی است که تغییرات جمعیت را با صنعتی شدن و رشد مدرنیته در یک جامعه نشان می‌دهد. این نظریه معمولاً شامل گذار از نرخ بالای زادوولد و مرگ‌ومیر به نرخ پایین تولد و مرگ است که با تغییرات در عوامل اجتماعی و اقتصادی و فرهنگی یک جامعه تفسیر می‌شود.

1. Fertility Rate
2. Life Expectancy
3. Population Pyramids
4. Population Growth and Decline
5. Demographic Transition Theory

۳- چیستی علم داده‌های جمعیتی

علم داده‌های جمعیتی به کاربرد ابزارها و تکنیک‌های علم داده برای تجزیه و تحلیل و درک مجموعه داده‌های قابل توجه در مورد جمعیت‌های انسانی اشاره دارد. این داده‌ها می‌تواند شامل داده‌های جمعیت‌شناختی (به عنوان مثال، سن، جنسیت، قومیت)، داده‌های بهداشتی (به عنوان مثال، شیوع بیماری، نرخ واکسیناسیون)، داده‌های اجتماعی-اقتصادی (به عنوان مثال، درآمد، سطح تحصیلات) و داده‌های رفتاری (مانند فعالیت رسانه‌های اجتماعی، الگوهای مهاجرت باشد) باشند. بدین جهت است که دولت‌ها و سازمان‌ها از داده‌های جمعیتی برای تدوین سیاست‌های مربوط به مراقبت‌های بهداشتی، آموزشی، تنظیم خانواده، رفاه اجتماعی، کار و امور اجتماعی و ... استفاده می‌کنند. این سیاست‌ها به عنوان سیاست‌ها و برنامه‌ریزی جمعیتی در یک کشور تدوین می‌شوند. برای رسیدن به هر نوع تصمیم‌سازی و تصمیم‌گیری در این سیاست‌ها و ایجاد امکان رسیدگی به چالش‌های جمعیتی و ترویج توسعه پایدار، تصمیم‌گیری مبتنی بر داده ضروری است.

کلان داده از اشکال سنتی داده‌های جمعیت‌شناختی و بهداشتی متمایز است زیرا به عنوان محصول جانبی فعالیت‌های انسانی تولید می‌شود. داده‌های سنتی به سرشماری[1]، بررسی‌ها، ثبت احوال، و انواع خدمات بهداشتی و سوابق اداری اشاره دارد که برای دهه‌ها و در برخی موارد قرن‌ها جمع‌آوری شده‌اند (Ashford et al., 2022). برای تجزیه و تحلیل و تفسیر داده‌های جمعیت، تکنیک‌های آماری و مدل‌سازی مبتنی بر علم داده با استفاده از ابزارهای کلان‌داده، داده‌کاوی، یادگیری ماشین و تجسم داده‌ها، نیاز است. علم داده‌های جمعیتی رشته‌های مختلفی از جمله آمار، علوم کامپیوتر، اپیدمیولوژی، علوم اجتماعی و بهداشت عمومی را در بر می‌گیرد. این علم از روش‌های محاسباتی و آماری پیشرفته برای تجزیه و تحلیل و تفسیر داده‌های جمع‌آوری‌شده از منابع مختلف، مانند نظرسنجی‌ها، سوابق اداری، پرونده الکترونیک سلامت، رسانه‌های اجتماعی و سایر منابع داده و کلان داده استفاده می‌کند.

هدف اصلی علم داده‌های جمعیتی دستیابی به درک عمیق‌تر از روندها، الگوها و ویژگی‌های جمعیت است. این کار به محققان اجازه می‌دهد تا تحولات پیش رو و قابل پیش‌بینی را شناسایی

1. Census

کنند، مداخلات را ارزیابی نمایند و تصمیمات آگاهانه‌ای برای بهبود سلامت عمومی، سیاست‌گذاری و مداخلات اجتماعی بگیرند.

مؤلفه‌های کلیدی علم داده‌های جمعیتی عبارت‌اند از:

- جمع‌آوری داده‌ها: جمع‌آوری داده‌ها از منابع مختلف، مانند سرشماری‌ها، نظرسنجی‌ها، سازمان‌های ثبت احوال و ثبت وقایع حیاتی و سازمان‌های بهداشت و درمان و سیستم عامل‌های دیجیتال انجام می‌شود. این منابع داده ممکن است شامل اطلاعاتی در مورد نرخ تولد، نرخ مرگ و میر، الگوهای مهاجرت، عوامل اجتماعی-اقتصادی و موارد دیگر شود.
- تحلیل و مدل‌سازی داده‌ها: استفاده از تکنیک‌های آماری، یادگیری ماشین و داده‌کاوی برای تجزیه و تحلیل و مدل‌سازی داده‌های مرتبط با جمعیت که می‌تواند شامل شناسایی روندها، الگوها، همبستگی‌ها[1] و پیش‌بینی پویایی‌های جمعیتی آینده شود.
- تفسیر داده‌ها: تفسیر نتایج تجزیه و تحلیل داده‌ها برای استخراج بینش‌های معنادار و نتیجه‌گیری در مورد ویژگی‌ها، رفتارها و روندهای جمعیت.
- سیاست‌گذاری و تصمیم‌گیری: استفاده از بینش داده‌های جمعیت برای اطلاع‌رسانی در مورد تصمیم‌گیری‌های خط مشی، تخصیص منابع و برنامه‌ریزی اجتماعی. برای مثال، دولت‌ها ممکن است سیاست‌های مراقبت‌های بهداشتی، آموزشی و برنامه‌ریزی شهری را بر اساس روندهای جمعیتی توسعه دهند.
- تجسم داده‌ها: ایجاد نمایش‌های بصری معنادار از داده‌های جمعیت برای تسهیل ارتباط و درک برای سیاست‌گذاران، محققان و عموم مردم.

همان‌طورکه روند گوگل در شکل (۱) نشان می‌دهد، علم داده‌های جمعیتی ابتدا در سال ۲۰۱۲ با میزان علاقه پایینی مطرح شد ولی مورد استقبال زیادی قرار نگرفت تا اینکه به طور جدی از سال ۲۰۱۵ همزمان با گسترش و توسعه علم داده طرح و در دو سال اخیر به کانون توجهات خصوصاً در جغرافیای ایالات متحده، کانادا و نیز اروپای غربی و استرالیا تبدیل شده است:

1. Correlations

شکل ۱. روندیابی علم داده‌های جمعیتی در گوگل؛ نمودار عمودی میزان علاقه به این واژه را از سال ۲۰۰۴ الی ۲۰۲۴ نشان می‌دهد. شکل سمت چپ بیانگر پراکندگی این علاقه در جغرافیای جهانی است.

(منبع: Google Trends 2024)

بنابراین، علم داده‌های جمعیتی با اتکا بر پیشرفت‌های علم داده و با تمرکز بر داده‌های جمعیتی تلاش دارد تا با استفاده از پیچیدگی‌ها و فناوری‌های متعدد حوزه داده و کلان‌داده، اقدام به ارائه الگوی معماری تجمیع‌شده‌ای از همه ظرفیت‌های داده‌محور حوزه جمعیت نماید بطوریکه مبتنی بر پیوند داده‌ها، ارائه زیرساخت و معماری، استفاده از هوش مصنوعی و سرویس‌های کیفی در تصمیم‌سازی مبتنی بر مقیاس بالای داده و در فضای قابل اشتراک ملی یا فراملی تحت پوشش سرویس رایانش ابری به توسعه فناورانه حوزه‌های تحلیل جمعیتی کمک نماید. این امکان در قالب داشبورد (یا پرتال) علم داده جمعیتی[1] قابلیت ارائه دارد. داشبوردها ابزارهایی هستند که به جمع‌آوری داده‌ها از منابع مختلف در زمان واقعی و به تصمیم‌گیری کمک می‌کنند. آنها ابزارهایی برای القای شفافیت در تصمیم‌گیری و توضیح تصمیم سیاست هستند (Head, 2020). هدف دیگر داشبوردها ایجاد آگاهی در مورد روند موارد، ردیابی و نظارت بر همه‌گیری‌ها و در عین حال هدایت اقدامات اصلاحی به موقع است (Mitra, 2021).

٤- چرایی علم داده‌های جمعیتی (کاربردها)

علم داده‌های جمعیتی به دلیل کاربردهای گسترده و نقش آن در ارائه بینش درباره جمعیت‌های انسانی از اهمیت قابل توجهی برخوردار است. چند دلیل کلیدی برای اهمیت و کاربردهای مختلف این علم عبارت‌اند از:

1 Population Data Science dashboard

- تصمیم‌گیری آگاهانه: علم داده‌های جمعیتی بینش‌های مهمی را ارائه می‌دهد که تصمیمات خط مشی، تخصیص منابع و برنامه‌ریزی اجتماعی را مشخص می‌کند. تصمیم‌گیرندگان می‌توانند از اطلاعات جمعیت شناختی برای طراحی برنامه‌ها و سیاست‌های موثری که نیازهای گروه‌های جمعیتی خاص را برآورده می‌کنند، استفاده کنند.
- بهداشت عمومی: علم داده‌های جمعیتی به ردیابی الگوهای بیماری، تجزیه و تحلیل نابرابری‌های بهداشتی و توسعه مداخلات بهداشت عمومی هدفمند کمک می‌کند. این امر به ویژه برای پیشگیری از بیماری، دسترسی به مراقبت‌های بهداشتی، و تحقیقات اپیدمیولوژیکی مهم است.
- برابری اجتماعی: با تجزیه و تحلیل داده‌های جمعیت، محققان و سیاست‌گذاران می‌توانند تفاوت‌ها را در آموزش، اشتغال و مراقبت‌های بهداشتی در بین گروه‌های جمعیتی مختلف شناسایی کنند. این اطلاعات برای ارتقای برابری اجتماعی و کاهش نابرابری‌ها بسیار مهم است.
- برنامه‌ریزی اقتصادی: داده‌های جمعیت به پیش‌بینی روند نیروی کار، رفتار مصرف‌کننده و رشد اقتصادی کمک می‌کند. کسب و کارها و دولت‌ها می‌توانند از این اطلاعات برای تصمیم‌گیری آگاهانه در مورد سرمایه‌گذاری، ایجاد شغل و سیاست‌های اقتصادی استفاده کنند.
- برنامه‌ریزی شهری: داده‌های جمعیتی برای طراحی شهرهایی که جمعیت رو به رشد را در خود جای دهند و در عین حال زیرساخت‌ها، مسکن و سیستم‌های حمل و نقل کارآمد را حفظ کنند، ضروری است.
- مهاجرت و اسکان مجدد: درک الگوهای مهاجرت برای مدیریت مهاجرت، اسکان مجدد پناهندگان و رسیدگی به مسائل مربوط به جابجایی و ادغام مهم است.
- پیری جمعیت: با افزایش سن جمعیت، داده‌های جمعیت شناختی به رفع چالش‌های مربوط به مراقبت از سالمندان، خدمات مراقبت‌های بهداشتی و برنامه‌ریزی بازنشستگی کمک می‌کند.
- برنامه‌ریزی مراقبت‌های بهداشتی: داده‌های جمعیت برای ارزیابی نیازهای مراقبت‌های بهداشتی، تخصیص منابع پزشکی و برنامه‌ریزی برای زیرساخت‌های مراقبت‌های بهداشتی

آینده استفاده می‌شود. این به پیش‌بینی شیوع بیماری، تجزیه و تحلیل استفاده از مراقبت‌های بهداشتی و اطمینان از دسترسی عادلانه به خدمات پزشکی کمک می‌کند.

- سیاست آموزشی: داده‌های جمعیت به مقامات آموزش و پرورش کمک می‌کند تا برای ظرفیت مدارس برنامه‌ریزی کنند، منابع را تخصیص دهند و برنامه‌های آموزشی را بر اساس نیازهای گروه‌های سنی و مناطق مختلف طراحی کنند.
- برنامه‌ریزی خانواده و سلامت باروری: داده‌های جمعیت با ارائه بینش‌هایی در مورد نرخ باروری، نرخ تولد و استفاده از روش‌های پیشگیری از بارداری، ابتکارات تنظیم خانواده را اطلاع می‌دهد. این امر به اطمینان از دسترسی به خدمات بهداشت باروری و آموزش تنظیم خانواده کمک می‌کند.
- مسکن و توسعه شهری: داده‌های جمعیتی برنامه‌ریزان شهری را در طراحی راه‌حل‌های مسکن، رسیدگی به کمبود مسکن و ایجاد جوامع پایدار و قابل زندگی راهنمایی می‌کند.
- پایداری محیطی: با درک پویایی جمعیت، سیاست‌گذاران می‌توانند استراتژی‌هایی را برای مدیریت منابع طبیعی، کاهش اثرات زیست‌محیطی و ترویج توسعه پایدار ایجاد کنند.
- نمایندگی سیاسی: داده‌های جمعیت بر تقسیم‌بندی مجدد سیاسی و تخصیص کرسی‌های قانونگذاری برای تضمین نمایندگی عادلانه برای جوامع مختلف تأثیر می‌گذارد.
- تحقیق بازار: کسب و کارها از داده‌های جمعیت برای هدف قرار دادن گروه‌های مصرف کننده خاص، توسعه محصولات و تنظیم استراتژی‌های بازاریابی بر اساس ویژگی‌های جمعیتی استفاده می‌کنند.
- تخصیص خدمات اجتماعی: سازمان‌های دولتی از داده‌های جمعیت‌شناختی برای تخصیص منابع برای خدمات اجتماعی مانند برنامه‌های رفاهی، مراقبت از سالمندان و حمایت از معلولان استفاده می‌کنند.
- مدیریت بلایا: داده‌های جمعیت به آمادگی و واکنش مسئولان در برابر بلایا کمک می‌کند تا تأثیر بالقوه بلایای طبیعی را بر جوامع آسیب‌دیده برآورد کنند.

یکی از جنبه‌های مهم داده‌های جمعیتی این اســـت که آنها عمدتاً برای تحقیق جمع‌آوری نمی‌شوند، بلکه بیشتر برای اهداف عملیاتی یا اداری جمع‌آوری می‌شوند (Jorm, 2015; Hand,

2018; McGrail et al., 2018; Keusch and Kreuter, 2021). در نتیجه، محققان کنترل بسیار کمتری بر چنین داده‌ها و نحوه پردازش آنها دارند، و تنها راه‌های محدودی برای یادگیری منشأ داده‌ها دارند، که مشخص نیست که آیا چنین داده‌هایی برای هدف یک مطالعه تحقیقاتی خاص مناسب هستند یا خیر (Biemer et al., 2017). در علم، سه چیز مهم است: داده‌ها، روش‌های مورد استفاده برای جمع‌آوری داده‌ها (که به آنها ارزش اثباتی می‌دهد)، و منطقی که داده‌ها و روش‌ها را به نتیجه‌گیری متصل می‌کند. وقتی هم داده‌ها و هم جمع آوری آنها خارج از کنترل یک محقق باشد، کار علمی مناسب می‌تواند چالش برانگیز باشد (Brown et al., 2018). بنابراین شکل‌گیری یک کلّ یکپارچه تحت عنوان علم داده‌های جمعیتی یک ضرورت اجتناب‌ناپذیر است.

۵- پیشینه موضوع

علم داده به همکاری چند رشته‌ای و تحقیق و توسعه مداوم برای ایجاد زیرساخت‌های علمی و همزمان استفاده از آن زیرساخت برای استخراج دانش جدید نیاز دارد. در واقع سرمایه‌گذاری‌های قابل توجهی در مراکزی صورت گرفته است که وظیفه نگهداری، ایجاد پیوستگی[1] و/یا یکپارچه‌سازی و در دسترس قراردادن داده‌ها برای تحقق این امر را بر عهده دارند. در حوزه جمعیت، تمرکز این سرمایه‌گذاری‌ها بر داده‌هایی است که عمدتاً بر روی افراد متمرکز و اغلب مبتنی بر جمعیت و مطالعات علمی مرتبط با آن هستند. نمونه‌های مهم اخیر از سرمایه‌گذاری‌ها در این حوزه را می‌توان به شرح ذیل برشمرد:

- **استرالیا**: در ارائه خدمات بهداشتی و انسانی، دولت استرالیا، سازمان‌های خصوصی و آژانس‌های غیرانتفاعی اطلاعات زیادی درباره استرالیایی‌ها جمع‌آوری می‌کنند. این اطلاعات یک منبع ملی ارزشمند است که می‌تواند برای بهبود درک از بیماری، توسعه درمان و بهبود خدمات استفاده شود. شبکه تحقیقات سلامت جمعیت[2] استرالیا یک همکاری ملی است که امکان تجمیع و ارائه داده‌های موجود از سراسر استرالیا را برای تحقیقات مهم فراهم می‌کند. این شبکه در سال ۲۰۰۹ توسط دفتر برنامه واقع در پرث، استرالیای غربی به‌عنوان یک شبکه ملی تأسیس و به‌عنوان برنامه ابتکاری دولت استرالیا در

1. Linkage
2. The Population Health Research Network (PHRN)

قالب استراتژی زیرساخت تحقیقات مشترک ملی[1]، اجرا شد. دولت‌ها و موسسات دانشگاهی، ایالتی و منطقه‌ای کمک‌های قابل توجهی به فعالیت‌های این شبکه از طریق واحدهای پیوند داده واقع در هر ایالت/ منطقه استرالیا کرده‌اند. پیوند داده‌های جمعیت برای بهبود رفاه همه استرالیایی‌ها مبتنی بر هدایت و فعال کردن پیوند همه داده‌ها برای تحقیقات اقدام محور در سطح جهانی از طریق همکاری مشترک برای فعال کردن زیرساخت‌های ارتباط داده‌ای نوآورانه و استفاده از آن برای حمایت از تحقیقات اقدام‌محور کلاس جهانی، هدف این شبکه است.[2]

- **نیوزیلند**: زیرساخت داده‌های یکپارچه[3] یک پایگاه داده تحقیقاتی بزرگ در نیوزیلند است که ریزداده‌های شناسایی نشده در مورد افراد و خانواده‌ها را در اختیار دارد. این داده‌ها در مورد رویدادهای زندگی مانند تحصیل، درآمد، مزایا، مهاجرت، عدالت و سلامت هستند که از سازمان‌های دولتی، نظرسنجی‌های ملی و سازمان‌های غیردولتی بدست می‌آید. داده‌ها به یکدیگر متصل شده و یکپارچه می‌شوند تا IDI را تشکیل دهند. IDI مکمل پایگاه داده کسب و کار طولی[4] (تجمیع کلیه داده‌های کسب و کار) است که ریزداده‌های مرتبط در مورد مشاغل را در اختیار دارد. این دو پایگاه داده از طریق داده‌های مالیاتی به هم مرتبط هستند. محققان از IDI برای انجام تحقیقات بین بخشی استفاده می‌کنند که بینشی را در مورد جامعه و اقتصاد فراهم می‌کند. این تحقیق می‌تواند به پاسخگویی به سوالاتی در مورد مسائل پیچیده‌ای که بر نیوزلندی‌ها تأثیر می‌گذارد کمک کند. هشت دسته کلی از داده‌ها در IDI عبارتند از: داده‌های بهداشتی، آموزش و پرورش، مزایا و خدمات اجتماعی، دادگستری، مردم و جوامع، جمعیت، درآمد و کار و مسکن.
- **کانادا**: محیط پیوستگی داده‌های اجتماعی آمار کانادا[6]، برنامه پیوند داده‌های اجتماعی در استفاده نوآورانه از داده‌های اداری و نظرسنجی‌های موجود برای رسیدگی به سؤالات مهم تحقیقاتی و اطلاع‌رسانی به سیاست‌های اجتماعی-اقتصادی از طریق پیوند سوابق را ترویج

1. National Collaborative Research Infrastructure Strategy framework
2. https://www.phrn.org.au
3. Integrated Data Infrastructure (IDI)
4. Longitudinal Business Database (LBD)
5. https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/
6. Social Data Linkage Environment (SDLE)

می‌دهد. SDLE پتانسیل یکپارچه‌سازی داده‌ها را در حوزه‌های مختلف، مانند سلامت، عدالت، آموزش و درآمد، از طریق ایجاد فایل‌های داده‌های تحلیلی مرتبط بدون نیاز به جمع‌آوری داده‌های اضافی از کانادایی‌ها، گسترش می‌دهد[1].

- **ایالات متحده** (شبکه ملی تحقیقات بالینی بیمار محور[2]): پیکورنت یک منبع ملی است که توسط PCORI تامین می‌شود، که در آن داده‌های بهداشتی با کیفیت بالا، مشارکت بیمار و تخصص تحقیقاتی پاسخ‌های سریع و قابل اعتمادی را ارائه می‌دهند. داده‌ها ستون فقرات پیکورنت هستند و مقیاس، کیفیت و امنیت داده‌های قابل دسترسی با پیکورنت یک عامل تمایز برای شبکه است. در این شبکه، داده‌های مربوط به برخوردهای روزمره با بیش از ۳۰ میلیون نفر در سراسر ایالات متحده را در هر سال نشان می‌دهد. از آنجایی که هر شریک شبکه پیکورنت تحت یک مدل داده مشترک[3] عمل می کند، کاربران می‌توانند به طور همزمان از میلیون ها بیمار سؤال مشابهی را بپرسند و پاسخ روشنی را حتی در سیستم های متفاوت دریافت کنند. نقطه قوت زیرساخت پیکورنت این است که محققان را قادر می‌سازد تا به سرعت بینش‌های داده‌ای را از شبکه بزرگی از سیستم‌های اطلاعات بالینی به‌طور هم‌زمان دریافت کنند، که از لحاظ تاریخی چالش‌برانگیز بوده است. یادگیری‌های مدیریت داده‌های پیکورنت در سایت گیت‌هاب بصورت مدل یادگیری متن باز و پیوسته به اشتراک گذاشته شده و برای همه قابل دسترسی است.[4]
- **بریتانیا** (موسسه فار، شبکه تحقیقات داده‌های اداری، و اخیراً تحقیقات داده‌های بهداشتی انگلستان[5]): هدف اصلی شبکه تحقیقاتی انفورماتیک سلامت بریتانیا[6] تقویت توانایی بریتانیا در تحقیقات انفورماتیک سلامت است. این امر با ایجاد مشارکت‌های فراگیر، ترویج هماهنگی فعالیت‌های کلیدی و عمل به عنوان نشانه‌ای از بهترین رویه هم در بریتانیا و هم در سطح بین‌المللی به این مهم دست خواهد یافت[7]. این شبکه توسط موسسه تحقیقات

1. https://www.statcan.gc.ca/en/sdle/index
2. The national patient-centered clinical research network, PCORnet
3. Common Data Model (CDM)
4. https://github.com/CDMFORUM
5. The Farr Institute, the Administrative Data Research Network, and most recently Health Data Research UK
6. UK Health Informatics Research Network ("Network")
7. https://gtr.ukri.org/projects?ref=MR%2FM501633%2F2

انفورماتیک بهداشتی فار که از چهار گره در سراسر بریتانیا توزیع شده و از دانشگاه کالج لندن، دانشگاه منچستر، دانشگاه سوانسی و دانشگاه داندی تشکیل شده است، رهبری خواهد شد. در مجموع موسسه فار شامل ۲۴ موسسه دانشگاهی بریتانیا و دو واحد با تخصص در تحقیقات انفورماتیک سلامت است. هدف آن ارائه تحقیقات پیشرفته و با کیفیت بالا است که داده‌های سلامت الکترونیکی را با سایر اشکال تحقیق و داده‌های جمع‌آوری شده به طور معمول پیوند می‌دهد و همچنین ظرفیت‌سازی در تحقیقات انفورماتیک سلامت را ایجاد می‌کند. این شامل همکارانی از همه بخش‌ها می‌شود که در تهیه، جمع‌آوری، تجزیه و تحلیل، مدیریت، کنترل و ارتقای داده‌های مربوط به سلامت و سلامت برای اهداف تحقیقاتی تأیید شده درگیر هستند. این افراد شامل، اما نه محدود به، دانشمندان کامپیوتر، مهندسان نرم‌افزار، آماردانان، ارائه‌دهندگان مراقبت‌های بهداشتی، پزشکان، متخصصان اخلاق، اطلاعات زیستی، دانشمندان علوم اجتماعی و تنظیم‌کننده‌ها می‌شود.

٦- چگونگی تحقق علم داده‌های جمعیتی

به منظور پیاده‌سازی علم داده‌های جمعیتی زمینه‌های انطباق با دنیای در حال تغییر شامل موضوعات ذیل را بایستی در نظر گرفت:

- **روش‌های پیوند داده**: پیشرفت‌های روش‌شناختی، از جمله توسعه نرم‌افزار و الگوریتم، تکنیک‌های افزایش حریم خصوصی، ارزیابی‌های کیفیت پیوند و استفاده‌های جدید از هوش مصنوعی در این بخش مدنظر قرار می‌گیرد.
- **زیرساخت داده**: پرداختن به چالش‌های پیوند داده‌های بین‌بخشی و بین‌حوزه‌ای، ایجاد سیستم‌های یکپارچه و متمرکز، برقراری مکانیسم‌های دسترسی طبقه‌بندی شده از جمله محصور نمودن داده‌ها، خودکارسازی ارزیابی‌های کیفیت داده، و ایجاد قابلیت همکاری با دیگران از موضوعات مورد بحث در بخش زیرساختی است.
- **تحقیق با استفاده از داده‌های دنیای واقعی**: استفاده نوآورانه از مجموعه داده‌های مرتبط و واقعی در این حوزه دارای اهمیت بسزایی است. تجزیه و تحلیل ویژگی‌های جمعیت‌شناختی، از جمله نژاد، قومیت، گرایش جنسی و هویت جنسی بسیار مهم است. ارائه

بینش‌های مربوط به تحقیقات پس از کووید-۱۹ و ارزیابی تبعات آن نیز از موارد جذاب داده‌های دنیای واقعی هستند.

- **پیامدهای اخلاقی، قانونی و اجتماعی**: تحقیق در مورد استفاده‌های مسئولانه و قانونی از داده‌های جمعیت، از جمله مشارکت و تسهیم عمومی، حاکمیت داده؛ جمع آوری و ترکیب بازخورد کاربران؛ ملاحظات هوش مصنوعی هنگام استفاده از داده‌های جمعیت؛ و سهیم نمودن جوامع عملی با لحاظ ملاحظات اخلاقی بسیار مهم است.

٦-١. پیوند داده‌های جمعیتی

پیوند داده، روشی برای گردآوری اطلاعات به دست آمده از منابع مختلف اما مربوط به یک فرد یا رویداد در یک فایل واحد است. پیوند داده یا رکورد به عنوان فرایند جفت کردن رکوردها از دو فایل و تلاش برای انتخاب جفت‌هایی که به یک موجودیت تعلق دارند، تعریف شده است (Winglee et. al, 2005). همچنین، پیوند داده‌ها به عنوان «به هم پیوستن دو یا چند منبع مختلف، داده هایی که به یک فرد، خانواده، مکان یا رویداد مربوط می شوند» تعریف شده است (Emery and Boyle, 2017). پتانسیل کامل داده‌های جمعیت اغلب تنها زمانی قابل دستیابی است که چنین داده‌هایی به پایگاه‌های داده دیگر مرتبط شوند. پیوند داده‌ها ارتباط اطلاعات افراد، مکان‌ها و رویدادها را امکان‌پذیر می‌کند. پیوند داده‌ها به در دسترس بودن شناسه‌های منحصر به فرد مشترک در میان مجموعه داده‌ها متکی است. شناسه‌های رایج مورد استفاده برای پیوند داده‌ها عبارت‌اند از: نام، آدرس، جنسیت، تاریخ تولد و تاریخ ثبت (Flack et al., 2019). سه گروه متمایز در فرآیند پیوند داده و دسترسی به آن وجود دارند[1]:

- متولیان داده‌ها[2]: متولیان یا نگهبانان داده‌ها افرادی هستند که جمع‌آوری داده‌ها را محقق و مدیریت می‌کنند. آنها در یک سازمان یا آژانس (مانند یک اداره بهداشت دولتی) کار می‌کنند و مسئول جمع‌آوری، استفاده و افشای امن داده‌ها هستند. متولیان داده‌ها اطلاعات شناسایی (مانند نام، جنسیت، تاریخ تولد) و همچنین اطلاعات محتوای مرتبط

1. https://www.phrn.org.au/about-us/data-linkage/whos-involved/
2. Data Custodian

(مانند اطلاعات بهداشـــتی مانند جزئیات تشـــخیص و درمان) را جمع‌آوری و ذخیره می‌کنند.

- پیونددهنده‌های داده[1]: پیوند دهنده‌های داده افرادی هســـتند که از اطلاعات شـــناســـایی برای ایجاد و حفظ شناسه‌های پیوند استفاده می‌کنند که اجازه می‌دهد داده‌ها در داخل و بین مجموعه‌های داده پیوند داده شوند.
- محققین: محققین افرادی هســـتند که از داده‌ها برای تجزیه و تحلیل و تحقیق اســـتفاده می‌کنند.

دو روش اصلی برای پیوند داده‌ها وجود دارد که عبارتند از: روش احتمالی[2] و روش قطعی[3]. پیوند احتمالی، پیوند رکوردها از دو (یا چند) فایل است و پیوند بر اساس احتمال توافق و عدم توافق بین طیف وسیعی از متغیرهای تطبیق‌یافته انجام می‌شود (Kelman et al., 2002). پیوند قطعی تطبیق دو رکورد بر اساس توافق دقیق شناسه‌های منحصر به فرد انجام می‌شود. این روش به یک شناسه منحصر به فرد یا تقریباً منحصر به فرد و داده با کیفیت بالا نیاز دارد (Kelman et al., 2010; Gill, 2001). به عنوان نمونه، بانک داده ایمنِ پیوند اطلاعات ناشناس[4] در سال ۲۰۰۶ در ولز انگلستان راه‌اندازی شده است. این بانک از توسعه ۶۴ پروژه تحقیقاتی حمایت کرده و در حال حاضر مجموعه‌ای از ۳۲۴ پروژه تحقیقاتی جاری را پشتیبانی می‌کند که موضوعات مختلفی را پوشش می‌دهد. بانک اطلاعات آن از محاسبات با کارایی بالا و بستر امن تحقیقات الکترونیکی[5] استفاده می‌کند تا به داده‌های پیچیده و چندگانه به صورت ایمن دسترسی داشته، مدیریت و تجزیه و تحلیل نماید. این بانک با انواع داده‌های نوظهور مانند متن آزاد، تصویربرداری، ژنتیک و داده‌های جغرافیایی کار می‌کند و پروژه‌های مربوط به هوش مصنوعی، یادگیری ماشین و پردازش زبان طبیعی را تسهیل می‌نماید. در این پروژه پیشرفته‌ترین مکانیسم پیوند داده‌ها پیاده شده و هر گونه بی‌نظمی را برجسته می‌کند و کیفیت و ثبات داده‌ها را برای تحقیق بهبود می‌بخشد.

1. Data Linkers
2. Probabilistic
3. Deterministic
4. Secure Anonymized Information Linkage (SAIL)
5. Secure e-Research Platform (SeRP)

٦-٢. داشبورد مدیریت داده‌های جمعیتی

طبق تعریف گارتنر، «داشبوردها مکانیزم گزارشی هستند که معیارها و شاخص‌های کلیدی را جمع‌آوری و نمایش می‌دهند تا بتوانند در یک نگاه توسط همه مخاطبان ممکن بررسی شوند. داشبوردها با آشکارسازی و انتقال بینش‌های متنی به معیارها برای نمایش شاخص‌ها با تجسم بصری، از جمله نمودارها، مقیاس‌ها، گیج‌ها و چراغ‌های راهنمایی که پیشرفت نشانگر را به سمت اهداف تعریف‌شده نشان می‌دهند، به بهبود تصمیم‌گیری کمک می‌کنند» (Gartner, 2021). مدیریت داده‌های جمعیتی باید تمام مراحل یک پروژه تحقیقاتی از جمله جمع‌آوری داده‌ها، ایجاد، دسترسی، استفاده، تجزیه و تحلیل، افشاء، ذخیره‌سازی، نگهداری، دفع، اشتراک‌گذاری و استفاده مجدد از آنها را پوشش دهد. یک برنامه مدیریت داده باید شامل: امنیت فیزیکی، سیاست‌ها و روندها، ترتیبات قراردادی و مجوز و موافقت‌نامه‌های محرمانه، آموزش برای تیم تحقیقاتی، فرمی که داده‌های جمعیتی در آن ذخیره خواهند شد، اهدافی که داده‌ها برای آنها استفاده و/یا افشا خواهند شد، شرایطی که تحت آن ممکن است دسترسی به داده‌های جمعیتی به دیگران اعطا شود و در نهایت اینکه چه اطلاعاتی از برنامه مدیریت داده‌های جمعیتی باید به شرکت‌کنندگان منتقل شود باشد (Flack et al., 2019):

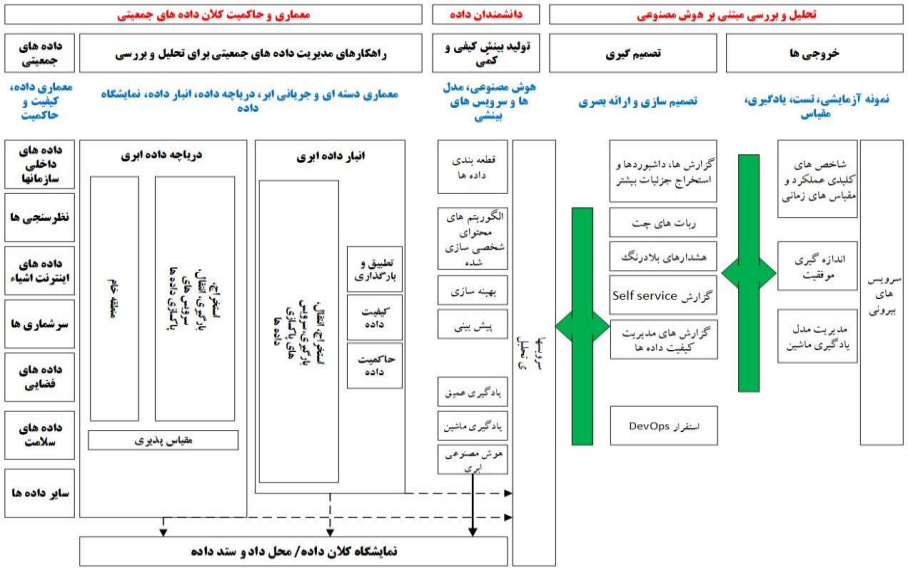
انواع داده‌هایی که اغلب به‌عنوان کلان‌داده دسته‌بندی می‌شوند، داده‌های دررو[1] (داده‌هایی که به‌طور تصادفی برای اهدافی غیرمرتبط با تحقیق ایجاد شده‌اند) هستند. به دلیل این محدودیت، جمعیت‌شناسان در سازمان ملل خواستار ایجاد یک اکوسیستم داده جدید شدند که فراتر از داده‌های دررو است و انواع داده‌های قابل تعمیم جمعیت را که مبنای تجزیه و تحلیل‌های جمعیتی خوب هستند، در بر می‌گیرد (HLG-PCCB, 2016). با افزایش حجم داده‌های جمعیتی، با انواع قالب‌ها (بدون ساختار[2]، نیمه ساختاریافته[3] و ساختاریافته[4]) که با سرعت بالا تولید می‌شوند و ناچار بایستی پردازش شوند، کشورها موظف به تحقق یک پارادایم کلان داده[5] چابک هستند (Moreno et al., 2019) که قادر به انطباق با این چالش‌های جدید باشد.

1. Eexhaust data
2. Unstructured
3. Semi-structured
4. Structured
5. Big Data paradigm

داشـــبوردها را با توجه به نوع‌شـــناســـی در ســـه دســـته‌بندی می‌توان قرار داد: داشـــبوردهای تصـمیم‌گیری، داشبوردهای ایستا برای اطلاع‌رسانی و داشبوردهایی برای ایجاد انگیزه و یادگیری (Sarikaya et al., 2019). داشـبوردهای تصـمیم‌گیری از تصـمیم‌گیری اسـتراتژیک یا عملیاتی پشتیبانی می‌کنند. به طور معمول، بسیار تعاملی هستند و سازمان‌ها را هدف قرار می‌دهند و به بیننده امکان سازگاری زیادی می‌دهند. به طور کلی، آنها امکان تجزیه و تحلیل معیار، شناسایی مناطق با عملکرد بالا/ پایین با تمرکز بر بهبود سـازمانی را می‌دهند. معمولاً به صـورت روزانه در زمینه‌های کسب‌وکار استفاده می‌شوند و ممکن است داده‌های زمان واقعی (سطح عملیاتی) یا داده‌های زمان متوسـط/ طولانی (سـطح اسـتراتژیک) را ارائه دهند. داشـبوردهای ایسـتا تمایل به عدم تعامل با کاربران خود دارند و معمولاً برای انتشـار دانش عمومی و آگاهی اسـتفاده می‌شـوند. داشـبوردهای سنتی‌تری هستند که در مدت زمان کوتاهی مصرف می‌شوند. داشبوردهای انگیزشی و یادگیری عموماً در خارج از محیط‌های سـازمانی یافت می‌شـوند و برای اهداف خاص یا کاربران عمومی بکار می‌روند (Beleza, 2022). توسعه داشبوردها در بخش عمومی اغلب مبتنی بر اهداف مختلفی مانند توانمندسازی مردم از طریق ایجاد شـفافیت و پاسـخگویی، نظارت بر عملکرد، گزارش‌دهی، برنامه‌ریزی، سـیاسـت‌گذاری و کاهش عدم تقارن اطلاعات اسـت (Matheus et al., 2020). براسـاس نظر بوهو، "جمعیت‌شـناسـان مدت‌هاسـت که داده‌های بزرگ را به روشـی کوچک جمع‌آوری و تجزیه و تحلیل کرده‌اند و تجزیه و تحلیل‌ها را محدود به آنچه سیستم‌های محاسباتی معمولی می‌توانند انجام دهند یا انتشـاردهندگان داده با دسـترسـی محدود اجازه خواهند داد، نموده‌انـد. برای بررسـی بهتر داده‌هـایی کـه در حـال حـاضـر در اختیـار داریم، بـایـد فرهنـگ جمعیت‌شـناسـی را تغییر دهیم تا ریزداده‌های جمعیتی را بزرگ بدانیم. این شـامل شـکل‌دهی به تعریف کلان‌داده جمعیتی، تغییر نحوه مفهوم‌سـازی مدل‌ها و ارزیابی مجدد نحوه ذخیره داده‌ها است" (Bohon, 2018).

در ادامه، یک چارچوب[1] رسـمی برای تحقق داشـبورد مدیریتی مبتنی بر علم داده‌های جمعیتی با تکیه بر استراتژی داده‌های بزرگ جمعیتی در فضای ابری با استفاده از الگوی ارائه شـده توسـط دکتر مورنو (Moreno et al., 2019) پیشنهاد می‌شود.

1. Platform



شکل ۲. معماری مفهومی چارچوب راهبردی پیاده‌سازی داشبورد علم داده‌های جمعیتی مبتنی بر کلان‌داده و رایانش ابری

٦-٢-١. کلان‌داده‌های جمعیتی

کلان‌داده‌های جمعیت شامل ریزداده‌های قابل تعمیم جمعیت است که می‌توان با الگوریتم‌های پیچیده‌ای که نیاز به محاسبات پیشرفته دارند به آنها نزدیک شد (Ruggles, 2014). داده‌های نشان داده شده در شکل ۲، شامل انواع داده‌های جمعیت به عنوان ورودی پلتفرم هوشمند داده‌های جمعیتی می‌شود:

- داده‌های داخلی سازمان‌ها: شامل همه انواع داده‌هایی می‌شود که در داخل هریک سازمان‌های مرتبط با موضوع جمعیت یا سایر سازمان‌ها ولی با موضوع جمعیت تولید می‌شود. به بیان دیگر این نوع داده‌ها دارای منشاء خارجی هستند. این می‌تواند شامل انواع تحلیل‌ها، پیش‌بینی‌ها، موارد بهره‌برداری از دانش نیروهای سازمانی تا انواع افزوده‌های داده‌ای کارشناسی در فرآیند جمع‌آوری در هر یک از سازمان‌ها باشد.
- داده‌های نظرسنجی به عنوان داده‌های جمعیتی به مجموعه‌ای از اطلاعات یا داده‌ها اطلاق می‌شود که در یک نظرسنجی از افراد جمع‌آوری شده است. این نوع داده‌ها

می‌توانند درباره نظرات، دلایل، خواسته‌ها یا سوالات اختصاصی باشند و برای تحلیل و درک نیازها و الگوهای جمعیتی مورد استفاده قرار گیرند. داده‌های نظرسنجی معمولاً از طریق پرسشنامه‌ها، مصاحبه‌ها، مطالعات میدانی، نظرسنجی‌های آنلاین یا روش‌های دیگری جمع‌آوری می‌شوند.

- داده‌های اینترنت اشیاء[1]: در لایهٔ معماری داده‌ها، یک منبع جدید اساسی برای داده‌های جمعیتی به‌عنوان اینترنت اشیا بایستی گنجانده شود. اینترنت اشیاء می‌تواند با اینترنت همه‌چیز[2] جایگزین شود. این یک پارادایم جدید فناوری است که شامل شبکه‌ای جهانی از ماشین‌ها و دستگاه‌هایی با قابلیت تعامل با یکدیگر، تصور می‌شود. اینترنت اشیا نقش مهمی در ایجاد جامعه هوشمند داشته و به‌عنوان عامل اصلی در نظارت هوشمند حوزه سلامت و جمعیت، یکی از اصلی‌ترین ورودی‌های داده در سال‌های آینده خواهد بود. ارزش پیشنهادی داده‌های اینترنت اشیا در سازمان‌های هوشمند هوشمند، دریافت داده‌های بلادرنگ است که مستقیماً از اشیاء متصل در اینترنت می‌آیند (Janssen and Kuk, 2016; Pan et al., 2016).
- داده‌های سرشماری از طریق فرآیند سرشماری جمع‌آوری می‌شوند. در سرشماری، جمعیت یک منطقه خاص یا یک کشور بررسی می‌شود و اطلاعات جامعی در مورد ویژگی‌های مختلف جمعیتی به دست می‌آید. این اطلاعات شامل تعداد افراد، همچنین مشخصات افراد مانند سن، جنسیت، تحصیلات، شغل و دیگر مواردی است که کمک می‌کند جامعه شناخت و تحلیل شود (Baffour et. al, 2013).
- داده‌های فضایی یا داده‌های جغرافیایی به عنوان داده‌های جمعیتی شامل اطلاعاتی هستند که مرتبط با مکان و موقعیت جغرافیایی افراد و جمعیت‌ها هستند. این داده‌ها می‌توانند از منابع متنوعی مانند سامانه‌های اطلاعات جغرافیایی، سنجش از دور، جی‌پی‌اس و دیگر روش‌ها جمع‌آوری شوند. با تلفیق داده‌های جمعیتی با داده‌های فضایی، اطلاعات مفیدی مانند تراکم جمعیت در مناطق مختلف، توزیع سنی و جنسیتی جغرافیایی، همبستگی‌های مکانی بین جمعیت و عوامل اقتصادی و سایر اطلاعات جمعیتی به دست می‌آید.

1. Internet of Things (IoT)
2. Internet of Everything (IoE)

- داده‌های سلامت به عنوان داده‌های جمعیتی شامل اطلاعات مرتبط با وضعیت سلامت جمعیت هستند. این داده‌ها شامل اطلاعاتی مانند بیماری‌ها، آمار مرگ و میر، شاخص‌های سلامت، عوامل ریسک، اعتیاد و سایر اطلاعات مربوط به سلامت فیزیکی و روانی جمعیت است. داده‌های سلامت جمعیتی می‌توانند از منابع مختلفی مانند دسترسی به پرونده‌های پزشکی، آمار بیمارستانی، سیستم‌های اطلاعات بهداشت عمومی، نظام اطلاعات بهداشتی یا سوابق فعالیت‌های بهداشتی جمعیتی جمع‌آوری شوند.

اگرچه در برخی موارد نظیر داده‌های داخلی سازمان‌ها یا داده‌های نظرسنجی و سرشماری، ممکن است در همان بدو امر، اطلاعات کلیدی بجای داده اولیه تولید شود ولی با توجه به نگاه معماری کلان‌نگر، همه انواع چنین ورودی‌هایی به مثابه داده و بعنوان فاز صفر یا ورودی چرخه مدیریت داده برای طراحی معماری مفهومی داده‌های جمعیتی در نظر گرفته شده است.

داده‌های جمعیت، همان‌طور که از پایگاه‌های اطلاعاتی اداری یا عملیاتی جمع‌آوری شده‌اند، درباره این هستند که مردم چه هستند و چه می‌کنند (Hand, 2018). این برخلاف داده‌های نظرسنجی است که معمولاً سؤالاتی درباره نگرش‌ها، باورها، انتظارات یا نیات با هدف درک رفتار افراد پرسیده می‌شود. اطلاعات واقعی در مورد افراد می‌تواند پاسخ‌های متفاوتی را در مقایسه با سؤالات مربوط به آنچه آنها ادعا می‌کنند ارائه کند، در حالی که استنباط باورهای افراد از رفتار آنها ممکن نیست (Christen and Schnell, 2023).

در فاز بعدی، نحوه مواجهه با این داده‌ها در ترکیب با هم برای ایجاد ظرفیت تحلیلی بیان می‌شود.

۲-۲-۶. راه‌کارهای مدیریت داده‌های جمعیتی برای تحلیل و بررسی

در گذشته، سازمان‌ها استراتژی‌های داده‌ای خود را از پایه می‌ساختند و معمولاً با یک انبار داده سنتی شروع می‌کردند. اما با افزایش حجم، مسیر و انواع داده‌ها، سازمان‌ها به دنبال راه‌حل‌های چابک‌تری برای هدایت استراتژی‌های تحلیلی خود هستند. یکی از این راه‌حل‌های چابک درنظر گرفته شده در معماری، لایه‌ای به نام راه‌کارهای مدیریت داده برای تحلیل و بررسی[1] به جای لایه

1. Data Management Solutions for Analytics (DMSA)

انبار داده[1] است. اگرچه مؤلفه انبار داده در حال حاضر برای بیشتر سازمان‌ها معتبر است، اما واضح است که حجم زیادی از داده‌ها، که مصداق کلان داده بر چنین داده‌هایی دلالت می‌کند، بی‌درنگ[2] و با فرمت‌های مختلف تولید می‌شوند. این موضوع با کاشت بیشتر حسگرهای اینترنت اشیا به مراتب روند صعودی به خود خواهد گرفت و لذا نمی‌توان آنها را در انبار داده معمولی با فرآیندهای مرسوم استخراج، انتقال و بارگذاری[3] مورد بحث قرار داد. طبق نامگذاری گارتنر، DMSA به عنوان سیستم نرم‌افزاری کاملی تعریف می‌شود که داده‌ها را در یک یا چند سیستم مدیریت فایل، معمولاً یک پایگاه داده یا چندین پایگاه داده، پشتیبانی و مدیریت می‌کند. این سیستم‌های مدیریتی شامل استراتژی‌های بهینه‌سازی خاصی هستند که برای پشتیبانی از پردازش تحلیلی طراحی شده‌اند که شامل: پردازش رابطه‌ای[4]، پردازش غیررابطه‌ای[5] (مانند پردازش گراف) و یادگیری ماشین، Json یا زبان‌های برنامه‌نویسی مانند Python یا R می‌شود و محدود به همین موارد نیست. داده‌ها لزوماً در یک ساختار رابطه‌ای ذخیره نمی‌شوند و می‌توانند از مدل‌های داده‌ای متعدد نظیر رابطه‌ای، XML و ... استفاده کنند" تعریف کرده است (Gartner report, 2019). گارتنر همچنین محصولات نرم‌افزاری مختلفی را که از DMSA پشتیبانی می‌کنند، ارزیابی می‌کند. این ارزیابی فقط محصولات اصلی را شامل می‌شود و بسیاری از آنها به طور خاص بر روی ابرهایی نظیر Google, Amazon AWS, Microsoft, Alibaba Cloud متمرکز هستند. در واقع، ابر اجازه می‌دهد تا این حجم بزرگ از داده‌ها را جذب[6] کند (Gartner report, 2019). DMSA سیستمی برای ذخیره[7]، دسترسی[8]، پردازش[9] و تحویل[10] داده‌های در نظر گرفته شده برای یکی از اصلی‌ترین موارد کاربردی مدنظر است که با این شیوه از تجزیه و تحلیل پشتیبانی می‌کند. این موارد کاربردی شامل پشتیبانی از انبار داده سنتی، عملیاتی، منطقی و مستقل از زمینه[11] است.

1. Data Warehouse (DW)
2. Real-time
3. Extract, Transform, Load (ETL)
4. Relational processing
5. Nonrelational processing
6. Ingestion
7. Storing
8. Accessing
9. Processing
10. Delivering
11. Context-independent

در حال حاضر، دو بخش اصلی در DMSA وجود دارد: دریاچه داده ابری[1] و انبار داده ابری[2] که مؤلفه ابری، مولفه اساسی هر دو خواهد بود. دریاچه داده یک مخزن داده متمرکز[3] است که اجازه می‌دهد حجم و تنوع بیشتری از داده‌ها را جمع‌آوری کرد و تمام داده‌های ساختاریافته، نیمه‌ساختاریافته و بدون ساختار را در هر مقیاسی بدون سختی و سربار معماری‌های انبار داده سنتی ذخیره می‌کند. اغلب اوقات این مخزن داده روی یک ارائه‌دهنده ابر اجرا می‌شود که در این صورت معمولاً دریاچه داده ابری نامیده می‌شود. دریاچه‌های داده دارای درجه بالایی از انعطاف‌پذیری و مقیاس‌پذیری هستند و می‌توان منابع داده متفاوت را در آنها وارد نمود؛ داده‌های با سرعت بالا را دریافت می‌کنند؛ می‌توانند در فرمت بومی ذخیره شوند؛ در مورد طرح داده اولیه نگرانی وجود ندارد؛ امکان اجرای پرس‌وجوهای SQL[4] موازی انبوه وجود دارد؛ و به الگوریتم‌های پیشرفته‌ای مانند یادگیری عمیق اجازه می‌دهد که تجزیه و تحلیل برای تصمیم‌گیری به‌لحظه را تقویت کند. ناحیه خام[5]، ظرفیت ذخیره‌سازی داده‌ها را دارد، یعنی داده‌های خام، بدون نیاز به ساختار اولیه این داده‌ها مطابق شکل ۲ قابل ورود به این سطح از دریاچه داده هستند. بنابراین، داده‌های خام، داده‌هایی هستند که در هنگام ذخیره‌سازی، دسته‌بندی[6] نشده‌اند. در نتیجه، وظایف آماده‌سازی[7]، پاکسازی[8] و تبدیل[9] داده‌ها حذف می‌شوند. معنای این فلسفه ذخیره‌سازی این است که: «ذخیره‌سازی داده‌ها به شکل اولیه‌اش ما را قادر می‌سازد تا از داده‌هایی که هنوز سؤالاتشان را نمی‌دانیم، پاسخ‌هایی پیدا کنیم. در حالی که یک انبار داده سنتی برای پاسخ به سوالاتی که قبلاً می‌دانیم بهینه شده است».

در یک محیط کلان داده، این ظرفیت اغلب به دلایل متعدد در یک انبار داده ابری قرار می‌گیرد. برای نمونه: کاهش هزینه، افزایش امنیت، ساده‌سازی تعمیر و نگهداری و امکان رشد نامحدود و آسان انبار داده. همانطور که گفته شد، داده‌ها به صورت تصاعدی در حال رشد هستند و سازمان‌ها باید با استفاده از معماری‌های مدرن انبار داده ابری، با این رشد روزافزون سازگار

1. Cloud Data Lake
2. Cloud Data Warehouse
3. Centralized data repository
4. Query
5. Raw Zone
6. Classify
7. Preparation
8. Cleansing
9. Transformation

شـوند. از دیگر مزایای اسـتفاده از ابر در این لایه این اسـت که سـازمان‌های هوشـمند می‌توانند زیرساخت‌های ابری خود را با استفاده از قابلیت مقیاس‌پذیری خودکار[1] برنامه‌ریزی نمایند. یکی از مزایای اصلی این کار، کاهش هزینه است. مقیاس‌پذیری خودکار به سرویس‌ها اجازه می‌دهد در زمان‌های کم‌باری به خواب بروند و در عین حال توانایی اداره پیک‌های ترافیک غیرمنتظره را داشته باشند. در فرآیندهای استخراج، بارگیری و انتقال که تحت عنوان ای‌تی‌ال[2] شناخته می‌شود، داده‌ها از منبع ورودی داده‌ها استخراج می‌شوند. ای‌ال‌تی متفاوت از از ای‌تی‌ال است. در ای‌تی‌ال، ابتدا داده‌ها از منابع مختلف استخراج می‌شوند. سپس، داده‌ها برای تحلیل و استفاده مناسب تبدیل می‌شوند. در نهایت، داده‌های تبدیل شده به منبع داده هدف/ هدف‌ها منتقل خواهند شد. در این روش پردازش، تمرکز بر تبدیل داده‌ها در مرحله میانی بین اسـتخراج و بارگیری اسـت. در ای‌ال‌تی[3]، داده‌ها ابتدا از منابع خارج، سـپس به سـامانه یا دریاچه داده مقصـد منتقل و در نهایت، تبدیل‌های لازم بر روی داده‌ها برای تحلیل و اسـتفاده صـورت می‌گیرد. در این روش، تمرکز بر تبدیل داده‌ها در مرحله پایانی پس از بارگیری است. انتخاب بین دو روش با توجه به نیازها و موارد کاربرد خاص و خامی داده انجام می‌شود.

۳-۲-٦. تولید بینش کیفی و کمی

در این مرحله با قطعه‌بندی داده‌ها، الگوریتم‌های مختلف بهینه‌سـازی، پیش‌بینی یا دسـته‌بندی مبتنی بر روش‌های مختلف آماری، یادگیری ماشـین، یادگیری عمیق و هوش مصـنوعی قابلیت پیاده‌سـازی دارند. این بخش مهم‌ترین وظیفه را در تولید بینش‌های کمّی و کیفی از داده‌های مختلف جمعیتی دارد و سـرویس‌های متنوع تحلیلی متناسـب با نیازهای مختلف در این بخش قرار می‌گیرند.

٤-۲-٤. تصمیم‌گیری

در این بخش، تصمیم‌سازی براساس بینش‌های ایجاد شده و ارائه‌های بصری لازم بر مبنای آنها شـکل می‌گیرد (Akhtar et al., 2020). گزارش‌های مختلف، داشـبوردهای مدیریتی و اجرایی، گزارش‌های خود_تولید شـده توسـط سـیسـتم، تولید و ارسـال هشـدارهای مختلف (بنابه تعاریف

1. Automatic scaling
2. ETL (Extract, Transform, Load)
3. ELT (Extract, Load, Transform)

گنجانده شده در سیستم نظیر اعلام هشدار کاهش جمعیت یا ...)، گزارش از کیفیت داده‌ها و میزان اعتماد به آنها و ... در این بخش انجام می‌شـــود. انواع ربات‌های چت برای ارتباط با مدیران یا مشـتریان نیز در این قسـمت قابل پیاده سـازی اسـت. هدف از اسـتقرار DevOps ، بهبود همکاری، اتوماسـیون و ارتباط بین تیم‌های توسـعه و عملیات به منظور تضـمین ارائه بهینه و قابل اعتماد راه‌حل‌های کلان داده اسـت. این قابلیت، شـامل مجموعه‌ای از روش‌ها و ابزارها اسـت که به سازمان‌ها کمک می‌کند فرایندهای توسعه نرم‌افزار و عملیات را بهبود بخشند.

۵-۲-٦. خروجی‌ها

در این قسمت ارتباط با سرویس‌های بیرونی، تبادل داده، مدیریت مدل‌های هوش مصنوعی و یادگیری ماشین، کیفیت‌سنجی میزان موفقیت معماری از طریق شاخص‌های مختلف سنجشی و کنترل کیفیت انجام خواهد شد.

معماری مفهومی ارائه شده به‌عنوان چارچوب مبتنی بر علم داده های جمعیتی از طریق داده‌های بزرگ، فضای ابری و یادگیری ماشینی با هدف جمع‌آوری و تجزیه و تحلیل مقادیر زیادی از داده‌ها برای به دست آوردن بینش در مورد جنبه‌های مختلف جمعیت شناختی یک جمعیت ارائه شد. این معماری روش‌های جمع‌آوری داده، زیرساخت محاسبات ابری و تکنیک‌های یادگیری ماشین را برای دستیابی به اهداف خود یکپارچه می‌کند. برای پیاده‌سازی این چارچوب ابتدا بایستی نسبت به تعیین پلتفرم رایانش ابری اقدام نمود. یک چارچوب رایانش ابری مناسب را بر اساس عواملی مانند مقیاس‌پذیری، قابلیت اطمینان، امنیت و مقرون به صرفه بودن انتخاب می‌کنند. برای اینکار گزینه‌هایی مانند خدمات وب آمازون (AWS)، Microsoft Azure یا Google Cloud Platform وجود دارد. برای ذخیره‌سازی و مدیریت داده‌های جمعیت می‌توان با استفاده از سرویس‌های ذخیره‌سازی ابری مانند Amazon S3، Azure Blob Storage یا Google Cloud Storage شروع به کار نمود. این سرویس‌ها قابلیت‌های مقیاس‌پذیر و بادوام ذخیره‌سازی را ارائه می‌دهند. در مرحله پردازش داده‌ها نیز یک خط لوله انتقال داده برای جمع‌آوری و پردازش داده‌های جمعیت با استفاده از ابزارهای پردازش داده مبتنی بر ابر مانند AWS Glue، Azure Data Factory یا Google Cloud Dataflow قابل پیاده‌سازی است. این خدمات امکان ادغام، تبدیل و بارگذاری داده‌ها از منابع مختلف را در فضای ذخیره‌سازی ابری فراهم می‌کنند. تامین امنیت و ایجاد حریم خصوصی داده‌ها از دیگر نکات مهم در پیاده‌سازی این معماری است. بنابراین بایستی اقدامات امنیتی مناسب را برای

محافظت از محرمانه بودن، یکپارچگی و در دسترس بودن داده‌های جمعیت اجرا کرد. این ممکن است شامل استفاده از مکانیسم‌های رمزگذاری، سیاست‌های کنترل دسترسی و استراتژی‌های پشتیبان‌گیری از داده‌ها باشد که توسط پلتفرم‌های پردازش ابری ارائه می‌شوند. برای استخراج بینش از داده‌های جمعیتی مبتنی بر یادگیری ماشین نیز می‌توان از خدماتی مانند AWS Sagemaker، Azure Machine Learning، یا Google Cloud AI Platform بهره برد. علاوه بر این، تکنیک‌های پردازش زبان طبیعی[1] را می‌توان برای تجزیه و تحلیل داده‌های ساختارنیافته مانند متن یا پاسخ‌های نظرسنجی به کار برد. اینکار امکان به دست آوردن بینشی در مورد احساسات، نظرات و ترجیحات عمومی را می‌دهد.

۷- نتیجه‌گیری

در این مقاله سعی شد تا ابتدا به چیستی علم داده‌های جمعیتی به عنوان یک حوزه میان رشته‌ای حاصل از تجمیع علم داده و علم جمعیت پرداخته شود و سپس چرایی و ضرورت آن تبیین گردد. در ادامه با ارائه مدل پیشنهادی داشبورد مدیریتی علم داده جمعیتی مبتنی بر کلان‌داده و رایانش ابری با تکیه بر دانش روزِ حوزه علم داده تلاش شد تا چگونگی این علم با معماری پیشرفته و چابکی از پیاده‌سازی علم داده‌های جمعیتی ارائه گردد. مدل ارائه شده این قابلیت را دارد تا ضمن اعمال حاکمیت موثرتر بر داده‌های جمعیتی به نحوی ارتباط بین داده‌های جمعیتی را برقرار نماید که توان تحلیل و بررسی، تولید بینش‌های کیفی، تصمیم‌سازی و ارائه سرویس‌های مختلف مبتنی بر هوش مصنوعی و یادگیری ماشین را بصورت یکپارچه و مبتنی بر اصول پیوند داده، در خود جای داده و در قالب داشبورد مدیریتی علم داده جمعیتی بصورت یک کلِّ جامع ارائه دهد. نگاه مبتنی بر علم داده برخلاف روش‌های محاسباتی و تکنیک‌های آماری مجزا و آفلاین ارائه شده در جوامع جمعیتی و مقالات مرسوم، درصدد است تا بر پایه قالبی پیشرفته و با استفاده از تنوع داده‌های آنلاین و آفلاین ورودی با تلفیق فناوری‌های مختلف یادگیری ماشین در فضای رایانش ابری مبتنی بر انبار داده ابری و دریاچه ابری اقدام به تعریف نوع جدید و پیشرفته‌ای از داشبوردها نماید که بالاترین ضریب بهره‌برداری از فناوری و عملی‌ترین حالت هوشمندی را در تحلیل و پردازش داده‌های جمعیت فراهم آورد.

1. Natural language processing (NLP)

فهرست منابع

Berntzen, L., Johannessen, M. R., & El-gazzar, R. “Smart Cities, Big Data and Smart Decision-making Understanding Big Data in Smart City Applications”, The Twelfth International Conference on Digital Society and EGovernments, c, 7–13, 2018.

Lori S. Ashford, Toshiko Kaneda, and Emmanuel Letouzé, “Demystifying Big Data for Demography and Global Health,” Population Bulletin 76, no. 1, 2022.

Subbotin, A., Aref, S. “Brain drains and brain gain in Russia: Analyzing international migration of researchers by discipline using Scopus bibliometric data 1996–2020,” Scientometrics 126, 7875–7900, 2021. (https://doi.org/10.1007/s11192-021-04091-x)

McGrail, K., Moran, R., O'Keefe, C., Preen, D., Quan, H., Sanmartin, C., Schull, M., Smith, M., Williams, C., Williamson, T., Wyper, G., Moorin, R., Kum, H.-C., Jones, K., Akbari, A., Bennett, T., Boyd, A., Carinci, F., Cui, X., Denaxas, S., Dougall, N., Ford, D., Kirby, R. S. and Kotelchuck, M. “A Position Statement on Population Data Science:: The science of data about people”, International Journal of Population Data Science, 3(1), 2018 (doi: 10.23889/ijpds.v3i1.415).

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. “What do we talk about when we talk about Dashboards?”, IEEE Transactions on Visualization and Computer Graphics, 3–22, 2019 (https://doi.org/10.5422/fso/9780823226351.003.0001).

Maciel, Frederico Brandão Beleza, “Development of a data report structure for Porto City Council: literature review and benchmark analysis on best dashboard design practices”, CPBS - Dissertações de Mestrado / Master Dissertations, 2022.

Matheus, R., Janssen, M., & Maheshwari, D. “Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities”, Government Information Quarterly, 37(3), 101284, 2020 (https://doi.org/10.1016/j.giq.2018.01.006).

Bohon, S.A. “Demography in the Big Data Revolution: Changing the Culture to Forge New Frontiers”, Popul Res Policy Rev 37, 323–341, 2018 (https://doi.org/10.1007/s11113-018-9464-6).

Head, M. G. “A real-time policy dashboard can aid global transparency in the response to coronavirus disease 2019. International Health, 12(5), 373–374, 2020 (https://doi.org/10.1093/inthealth/ihaa039).

Arun Mitra, Biju Soman, Gurpreet Singh, “An Interactive Dashboard for Real-Time Analytics and Monitoring of COVID-19 Outbreak in India: A proof of Concept”, In proceedings of the 1st Virtual Conference on Implications of Information and Digital Technologies for Development, 2021 (arXiv:2108.09937).

Louisa Jorm. “Routinely collected data as a strategic resource for research: priorities for methods and workforce”, Public Health Research Practice, 25 (4), 2015 (https://doi.org/10.17061/phrp2541540).

Janssen, M., & Kuk, G. “Big and Open Linked Data (BOLD) in research, policy, and practice”, Journal of Organizational Computing and Electronic Commerce, 26(1–2), 3–13, 2016 (https://doi.org/10.1080/10919392.2015.1124005)

Pan, Y., Tian, Y., Liu, X., Gu, D., & Hua, G. “Urban Big Data and the Development of City Intelligence”, Engineering, 2(2), 171–178, 2016 (https://doi.org/10.1016/J.ENG.2016.02.003)

David J Hand “Statistical challenges of administrative and transaction data. Journal of the Royal Statistical Society: Series A (Statistics in Society)”, 181 (3):555–605, 2018 (https://doi.org/10.1111/rssa.12315)

Kimberlyn M McGrail, Kerina Jones, Ashley Akbari, Tellen D Bennett, Andy Boyd, et al. “A position statement on population data science: The science of data about people”, International Journal of Population Data Science, 3(1), 2018 (https://doi.org/10.23889/ijpds.v3i1.415)

Florian Keusch and Frauke Kreuter “Digital trace data: Modes of data collection, applications, and errors at a glance”, In Handbook of Computational Social Science, Vol 1, pages 100–118. Taylor and Francis, 2021 (https://doi.org/10.4324/9781003024583-8)

Paul Biemer. Errors and inference. In: Ian Foster, Rayid Ghani, Ron S Jarmin, Frauke Kreuter, and Julia Lane, editors, “Big Data and Social Science”, chapter 10, pages 265–297. CRC Press, Boca Raton, 2017 (https://doi.org/10.1201/9781315368238).

Andrew W Brown, Kathryn A Kaiser, and David B Allison. “Issues with data and analyses: Errors, underlying themes, and potential solutions”, Proceedings of the National Academy of Sciences, 115(11):2563–2570, 2018 (https://doi.org/10.1073/pnas.1708279115).

Gartner, Gartner- Dashboard Definition, 2021 (https://www.gartner.com/en/information-technology/glossary/dashboard).

Winglee M, Valliant R, Scheuren F: “A case study in record linkage”, Survey Methodology, 31 (1): 3-11, 2005.

Emery, J., & Boyle, D. “Data linkage. Australian Family Physician”, 46(8), 615–619, 2017 (https://search.informit.org/doi/10.3316/informit.985328596864810).

F. Flack, A. Kemp-Casey, N. Wray, “using linked administrative data in clinical trials: A Guide for Clinical Trialists and Researchers”, Australian Clinical Trials Alliance, Retrieved from the ACTA website www.clinicaltrialsalliance.org.au, 2019.

Kelman CW, Bass AJ, Holman CD. “Research Use of Linked Health Data-A Best Practice Protocol”, Aust N Z J Public Health, 26(3):251-5, 2002.

Karmel R, Anderson P, Gibson D, Peut A, Duckett S, Wells Y. “Empirical Aspects of Record Linkage Across Multiple Data Sets Using Statistical Linkage Keys: The Experience of the PIAC Cohort Study”, BMC Health Serv Res, 2010.

Gill L. “Methods for Automatic Record Matching and Linking and Their Use in National Statistics”, Statistics N, editor. London, 2001.

HLG-PCCB (High-level group for partnership, coordination and capacity-building for statistics for the 20130 agenda for sustainable development). Global action plan for sustainable development data. Report, 2016 (https://unstats.un.org/sdgs/files/global-consultation-hlg-1/GAP_HLG-20161021.pdf).

Caio Moreno, Ramón Alberto Carrasco, Enrique Herrera-Viedma, “Data and Artificial Intelligence Strategy: A Conceptual Enterprise Big Data Cloud Architecture to Enable Market-Oriented Organizations”, International Journal of Interactive Multimedia and Artificial Intelligence, 2019.

Ruggles, S. “Big microdata for population research. Demography”, 51(1), 287–297, 2014.

Baffour, B., King, T. and Valente, P. “The Modern Census: Evolution, Examples and Evaluation”, International Statistical Review, 81: 407-425, 2013 (https://doi.org/10.1111/insr.12036).

Peter Christen, and Rainer Schnell, “Thirty-three myths and misconceptions about population data: from data capture and processing to linkage”, International Journal of Population Data Science, 2023.

Gartner: Data Management Solutions for Analytics, 2019 (https://www.gartner.com/reviews/customers-choice/datawarehouse-solutio).

Gartner: Magic Quadrant for Data Management Solutions for Analytics, 2019 (https://www.gartner.com/doc/3898487/magicquadrant-data-management-solutions).

Nikhat Akhtar, Nazia Tabassum, Asif Perwej, Yusuf Perwej “Data analytics and visualization using Tableau utilitarian for COVID-19”, Global Journal of Engineering and Technology Advances, 2020 (https://doi.org/10.30574/gjeta.2020.3.2.0029).