

کاوش روابط پنهان میان شاخص‌های چندگانه‌ی محرومیت با رویکرد ترکیبی داده‌کاوی و شبکه‌های پیچیده

احسان مؤمنی (کارشناس ارشد)

الهام آخوندزاده نوقایی* (دربی)

دانشکده‌ی مهندسی صنایع، مؤسسه‌ی آموزش عالی ابوتکی

بهروز مینایی بیدگلی (استادیار)

دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت

مهندسی صنایع و مدیریت شریف، زمستان ۱۳۹۴ (۱۳۹۴)
دربی ۱ - ۳۱، شماره‌ی ۲/۲، ص. ۱۳۱-۱۴۲، (پادداشت‌نوی)

بحث محرومیت‌زدایی از مناطق محروم در میان سایر برنامه‌های کشورها همواره از اهمیت ویژه‌ی برخوردار است. دولت‌ها می‌کوشند تا با درک نیازها و کمبودهای هر منطقه، اقداماتی در جهت رفع نیازهای اساسی مناطق انجام دهند. در این نوشتار سعی بر آن است تا با کشف ارتباطات پنهان بین شاخص‌های محرومیت و ارائه‌ی نتایج، گامی نوین در این حوزه برداشته شود. بدین‌منظور رویکرد جدیدی از کاربرد ابزارهای داده‌کاوی و شبکه‌های پیچیده در این حوزه ارائه شده که ترکیبی از تکنیک قوانین انجمنی و مفاهیم ساختاری شبکه‌های پیچیده است. در این رویکرد ابتدا با استفاده از قوانین انجمنی روابط پنهان بین شاخص‌های نه‌گانه‌ی محرومیت به‌دست آمده و در ادامه شبکه‌ی ابتکاری مبتنی بر قوانین، به‌منظور شناسایی تأثیرگذارترین و تأثیرپذیرترین شاخص‌ها مورد بررسی قرار گرفته است. از نتایج به‌دست آمده می‌توان در تصمیم‌گیری‌های محرومیت‌زدایی، اختصاص بودجه برای اجرای بهتر و تمرکزگرایی هدف‌مند در محرومیت‌زدایی مناطق استفاده کرد.

واژگان کلیدی: شاخص‌های چندگانه‌ی محرومیت، داده‌کاوی، قوانین انجمنی، شبکه‌های پیچیده.

۱. مقدمه

از مناطق محروم، می‌تواند نقش بسیار شگرفی در تعیین هدف‌مند سیاست‌ها و اولویت‌های برنامه‌ریزی در این حوزه داشته باشد.

با توجه به اهمیت موضوع و اهمیت محرومیت‌زدایی هدف‌مند و تمرکزگرایی در انجام آن، علاوه بر شناخت متغیرهای مهم و تأثیرگذار در محرومیت یک منطقه، کشف روابط پنهان در میان این متغیرها و شاخص‌ها در محروم شدن یک منطقه، می‌تواند کمک تأثیرگذاری را در ارائه‌ی راهبردها و سیاست‌هایی در جهت محرومیت‌زدایی هدف‌مند داشته باشد. این تأثیرگذاری‌ها در گذشته به‌صورت تجربی توسط کارشناسان مطرح می‌شد و در کشورها مورد استفاده قرار می‌گرفت و اکنون نیز مورد استفاده قرار می‌گیرد. به‌عنوان مثال در کشورهای توسعه‌یافته نظیر ژاپن، استرالیا و کشورهای اسکاندیناوی، دولت به‌جای افزایش حمایت مالی بخش درمان، بودجه‌ی مدیریت شهری را افزایش داده تا با بهبود کیفیت محیط زندگی افراد، وضعیت عمومی شهروندان نیز بهبود یابد.^[۱]

براساس بررسی‌های انجام‌شده، تاکنون تحقیقات مختلفی در زمینه‌ی کاربرد روش‌های آماری در شاخص‌های محرومیت، و تجزیه و تحلیل آن‌ها صورت گرفته و گزارش‌هایی نیز از وضعیت توسعه‌یافتگی و محرومیت مناطق ارائه شده است. در حوزه‌ی داده‌کاوی و در بخش توصیفی آن نیز با استفاده از روش‌های خوشه‌بندی نسبت به تحلیل خوشه‌ی داده‌های محرومیت مناطق براساس معیارها و شاخص‌های

در تمامی کشورها برخی از مناطق وجود دارند که در آن‌ها به‌علت نبود امکانات و تسهیلات یا دشواری دسترسی به آن‌ها از یک سو، و دورافتادگی‌های جغرافیایی و خرابی‌های ناشی از جنگ و غیره از سوی دیگر، سطح زندگی با کمبودهایی مواجه می‌شود و وضعیت معیشت زندگی دچار نقص‌هایی می‌شود. یکی از علل اصلی این امر عدم توجه به روابط پنهان شاخص‌های مختلف محرومیت، و تأثیرات نهانی آن‌ها بر یکدیگر است. در واقع درک درست از نیازهای اقشار مختلف جامعه در مناطق مختلف جغرافیایی و امکانات و تسهیلات در دسترس آن‌ها، و نیز کشف چگونگی تمرکزگرایی در محرومیت‌زدایی مناطق براساس روابط پنهان کشف شده، در حل این موضوع مهم نقش به‌سزایی دارد.

تفاوت‌های عمده و متعدد در سطح توسعه‌یافتگی و محرومیت در مناطق مختلف جامعه، در طولانی‌مدت می‌تواند آثار مخربی بر کارایی اقتصادی، سیاسی و اجتماعی کشورها داشته باشد. برای کاهش این نابرابری‌ها، شناسایی تفاوت‌های منطقه‌ی یک ضرورت است.^[۱] بنابراین در زمینه‌ی محرومیت‌زدایی از مناطق مختلف، توجه ویژه به متغیرهای تأثیرگذار در زندگی اقشار جامعه و تمرکزگرایی در جهت محرومیت‌زدایی

* نویسنده مسئول

تاریخ: دریافت ۱۳۹۲/۱۲/۱۱، اصلاحیه ۱۳۹۳/۳/۲۴، پذیرش ۱۳۹۳/۵/۱۳.

ehsanmomeni1989@yahoo.com
e.akhondzade.n@gmail.com
minaei@gmail.com

مختلف محرومیت اقدام شده است. اما از الگوریتم‌های کشف توالی و همچنین روش‌های پیش‌بینانه‌ی داده‌کاوی در این حوزه استفاده‌ی چندانی نشده است. در مطالعات آماری صورت گرفته در این زمینه، می‌توان به موردی در سال ۱۳۸۸ اشاره کرد^[۱] که در آن توسعه‌ی شاخص‌های کیفیت زندگی شهری در قالب مطالعه‌ی موردی در شهر نورآباد استان لرستان بررسی شده است. طی تحقیقات آماری و میدانی انجام گرفته در تحقیق مذکور براساس شاخص‌های تعریف شده، ۲۵ درصد خانوارها در این شهر در حالت بهزیستی، ۳۰/۱ درصد در حالت محرومیت، ۲۴/۴ درصد در حالت ناهماهنگی و ۲۰/۴ درصد در حالت انطباقی قرار گرفته‌اند. بر این اساس این شهر در زمره‌ی شهرهای محروم کشور قرار می‌گیرد؛ این نسبت در اغلب شهرستان‌ها و شهرهای کشور حاکم است. آن‌ها با استفاده از اطلاعات به دست آمده از این تحقیق، رویکردی در جهت تمرکزگرایی در محرومیت‌زدایی براساس نتایج آماری استخراج شده ارائه کرده‌اند.

در ارزیابی درجه‌ی محرومیت و توسعه‌یافتگی مناطق براساس شاخص‌های محرومیت و شاخص‌های توسعه‌یافتگی، روش‌های آماری مختلفی به کارگرفته می‌شود؛ نوع دیگری از این روش در پایان‌نامه‌ی با همین موضوع (۱۳۸۹) ارائه شده است.^[۲] در این تحقیق با بررسی و اندازه‌گیری فقر در شهر تهران -- با استفاده از روش‌های میانه و میانگین -- مخارج خوراکی و معکوس ضریب انگل و حداقل کالری مورد نیاز برای مناطق مختلف شهر تهران مورد اندازه‌گیری قرار گرفته است. با توجه به نتایج حاصل از این تحقیق، شهر تهران در مقایسه با شهرهای مختلف کشور از وضعیت نسبتاً بهتری برخوردار است. اما در داخل شهر تهران نیز مناطقی وجود دارد که براساس برخی شاخص‌ها وضعیت نسبتاً نامطلوبی دارند. در این نوع تحقیقات تنها گزارشات آماری در رابطه با این شاخص‌ها ارائه می‌شوند، اما استفاده از فنون داده‌کاوی و کشف روابط بوجود آورنده این محرومیت‌ها در مناطق شهری می‌تواند اهمیت ویژه‌ی داشته باشد که در این نوع تحقیقات استفاده نشده است.

از سوی دیگر مطالعاتی که از فنون داده‌کاوی در این حوزه استفاده کرده‌اند، بیشتر بر روش‌های توصیفی داده‌کاوی متمرکزند که غالباً از روش‌ها و الگوریتم‌های مختلف خوشه‌بندی بهره می‌گیرند. بخش‌بندی جغرافیایی مناطق براساس خوشه‌های به دست آمده روشی است که در اکثر مطالعات مربوط به شاخص‌های محرومیت مورد استفاده قرار می‌گیرد. یکی از اولین و کامل‌ترین مطالعاتی که با استفاده از این روش انجام شده^[۳] بررسی پایگاه داده‌ی شاخص‌های چندگانه‌ی محرومیت^۱ بوده است که در آن، با در نظر گرفتن همه‌ی شاخص‌ها و تمرکز بر شاخص سلامت و پراکندگی جغرافیایی جمعیت، این داده‌ها تجزیه و تحلیل شده است. در مطالعه‌ی یادشده با تعیین معیارهایی از میزان سلامت عمومی مناطق مختلف، به بخش‌بندی مناطق و ارائه‌ی گزارشی از وضعیت سلامت عمومی حاکم بر این مناطق پرداخته‌اند. در تحقیق دیگری (۲۰۱۱) تفاوت بین محرومیت شهری و روستایی براساس کشف مهم‌ترین شاخص‌ها در خوشه‌بندی داده‌ها و امتیازدهی به تمامی شاخص‌های محرومیت، تأثیر محرومیت هر خوشه بر خوشه‌های دیگر مورد بررسی قرار گرفت.^[۴] در سال ۲۰۱۰ نیز یک مدل همبستگی بین شاخص‌های محرومیت، جنسیت و درآمد به همراه شاخص سلامت و نرخ مراجعه به بیمارستان پیشنهاد شد.^[۵] با توجه به نتایج پیاده‌سازی این مدل بر پایگاه داده‌ی شاخص‌های چندگانه‌ی محرومیت سال ۲۰۰۷ یک همبستگی مثبت بین نرخ وقوع بیماری و درآمد منطقه‌های مختلف برقرار بود. در سال ۲۰۰۶ با تجزیه و تحلیل خوشه‌ی شاخص‌های محرومیت کشور سوییس، فاکتورهایی براساس توابع تولیدشده‌ی مبتنی بر خوشه‌های به دست آمده کشف^[۶] و برای رفع محرومیت مناطق مختلف آن

کشور یک استراتژی براساس خوشه‌های دندانه‌ی ارائه شد. در تحقیق دیگری که در سال ۲۰۱۱ انجام گرفت، روابط و همبستگی بین محرومیت‌های منطقه‌ی و ریسک سرطان‌های اتفاقی افتاده در مناطق مختلف کشور انگلستان مورد بررسی قرار گرفت.^[۷] براساس نتایج به دست آمده، با توجه به تعداد بسیار زیاد این وابستگی‌ها و داشتن معیارهای عددی از نظر همبستگی بین این دو متغیر، می‌توان گفت بین مناطق مختلف انگلستان و این پایگاه داده به‌وضوح ارتباط مستقیم برقرار است.

در بین تحقیقات انجام گرفته، تحقیقی درخصوص کشف روابط پنهان بین شاخص‌های محرومیت و کاربرد این شاخص‌ها در محرومیت‌زدایی مناطق انجام نشده است.

در تحقیق حاضر ضمن بررسی شاخص‌های چندگانه‌ی محرومیت -- یکی از کامل‌ترین شاخص‌ها در این حوزه -- به بررسی علمی این شاخص‌ها در بین مناطق، براساس روابط باهمایی پنهان آن‌ها می‌پردازیم. همچنین در رویکردی نوین با استفاده از مفاهیم شبکه‌های پیچیده به کشف مهم‌ترین شاخص‌ها و مناطق می‌پردازیم و بر این اساس راهکارهایی برای تمرکز بر محرومیت‌زدایی مناطق و هرچه بهتر انجام شدن این موضوع ارائه می‌دهیم. در واقع در این مقاله رویکرد ترکیبی جدیدی مبتنی بر فنون داده‌کاوی و شبکه‌های پیچیده به کارگرفته شده است. در این راستا ابتدا با اجرای تکنیک قوانین انجمنی روی داده‌های آماده‌سازی شده از شاخص‌های محرومیت، به کشف روابط پنهان میان این شاخص‌ها می‌پردازیم و از بین قوانین کشف شده، قوانینی که به لحاظ جذابیت و کاربرد مناسب‌اند استخراج می‌شوند؛ همچنین مفاهیم نهفته در قوانین تفسیر و پیشنهاداتی برای اجرای هرچه بهتر محرومیت‌زدایی مطرح می‌شود. سپس براساس قوانین به دست آمده، شبکه‌ی قوانین محرومیت تشکیل می‌شود و بعد از تجزیه و تحلیل شبکه، نتایج و مفاهیمی برای تمرکز در محرومیت‌زدایی مطرح خواهد شد. همچنین در این مقاله با توجه به نوع داده‌های شاخص‌های چندگانه‌ی محرومیت و به‌منظور استخراج نتایج بهتر، یک الگوریتم ابتکاری جدید برای گسسته‌سازی و آماده‌سازی داده‌ها، طراحی و مورد استفاده قرار گرفته است.

تحقیقات انجام شده در این زمینه بیشتر بر روش‌های آماری و بررسی همبستگی بین شاخص‌ها متمرکز بوده است. در این مقاله رویکرد جدیدی مبتنی بر داده‌کاوی و شبکه‌های پیچیده ارائه می‌شود که در مرحله‌ی اول ارتباط بین شاخص‌های مختلف را به صورت دقیق‌تر در قالب قوانین «اگر... آنگاه» نشان می‌دهد که این قوانین به‌آسانی قابل درک و استفاده‌اند. همچنین در رویکردی نوین، تأثیرگذاری شاخص‌های محرومیت بر یکدیگر، براساس قوانین به دست آمده و تحلیل شبکه‌های پیچیده بررسی می‌شود. با توجه به دقیق بودن این روابط و همچنین سهولت درک آن‌ها، تصمیم‌گیرندگان در این حوزه می‌توانند تصمیم‌هایی هدف‌مندتر و دقیق‌تر در حوزه‌ی محرومیت‌زدایی از مناطق محروم و ارتقای این شاخص‌ها و بهبود کیفیت زندگی شهروندان داشته باشند.

در بخش بعدی مقاله تعریف مسئله و رویکرد به کارگرفته شده در اجرای این تحقیق ارائه خواهد شد. در بخش سوم مرور ادبیات قوانین انجمنی و الگوریتم ابروری که در این تحقیق از آن استفاده شده است در حوزه‌ی محرومیت‌زدایی و تمرکز بر شاخص‌های محرومیت ارائه خواهند شد. در بخش چهارم چارچوب پیشنهادی برای اجرای تحقیق آورده خواهد شد و در بخش پنجم چگونگی پیاده‌سازی چارچوب پیشنهادی روی داده‌ها توضیح داده شده و ضمن استخراج نتایج، تشریح تفسیر آن‌ها ارائه خواهد شد. در بخش پایانی نیز نتایج نهایی به دست آمده و نکات مورد توجه ارائه می‌شود.

۲. ادبیات تحقیق

۱.۲. داده‌کاوی

در سازمان‌ها و شرکت‌های مختلف همواره داده‌ها تولید، ذخیره و در نهایت بازایی و گزارش‌گیری می‌شود. در میان همین انبوه داده‌ها ممکن است دانش‌های بسیار پرکاربردی وجود داشته باشد؛ جست‌وجوی این دانش‌ها در میان داده‌ها موضوعی است که ضرورت استفاده از علم داده‌کاوی را بیش از پیش نمایان می‌سازد. در واقع هدف داده‌کاوی ایجاد مدل‌هایی برای تصمیم‌گیری و تحلیل داده‌ها براساس اهداف مورد نظر است. متدولوژی کلی عملیات داده‌کاوی عبارت است از:

۱. انتخاب داده‌های مناسب از میان پایگاه داده،
۲. تبدیل داده‌ها به صورت مورد نیاز برای مدل‌سازی،
۳. کاوش و جست‌وجو برای کشف الگوها، نتایج و دانش‌های پنهان،
۴. تفسیر، تجزیه و تحلیل نتایج.^[۱۹]

۲.۲. قوانین انجمنی

استخراج قوانین انجمنی نوعی عملیات داده‌کاوی است که ارتباط بین ویژگی‌ها در پایگاه داده را جست‌وجو می‌کند. قوانین انجمنی جزء روش‌های غیرنظارتی و توصیفی داده‌کاوی است. نتایج این روش الگوها و قوانینی هستند که به صورت آگرو آنگاه ارائه خواهند شد. این قوانین با معیارهای پرکاربرد «پشتیبان^۲» و «اطمینان^۳» که بیان‌گر میزان کیفیت و عمومیت این قوانین هستند معرفی می‌شوند؛ پشتیبان بیان‌گر درصد نسبت مجموعه داده‌های سمت راست و چپ قانون به تعداد کل داده‌هاست، و اطمینان بیان‌گر میزان وابستگی سمت چپ و راست قانون به یکدیگر است.^[۱۹]

$A \rightarrow B$ (Support, Confidence)

$$\text{Support}(A, B) = \frac{\text{Count}(A, B)}{\text{The Total Number Of Records}}$$

$$\text{Confidence}(B|A) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$$

۱.۲.۲. الگوریتم اپریوری یا پیشینار

الگوریتم اپریوری یکی از قوی‌ترین و کامل‌ترین الگوریتم‌های کشف قوانین انجمنی است که با صرفه‌جویی در زمان و حافظه‌ی مورد نیاز برای تجزیه و تحلیل، به جست‌وجو برای یافتن اقلام و زیرمجموعه‌های پرتکرار در پایگاه داده می‌پردازد. الگوریتم اپریوری از دو مرحله‌ی کلی تشکیل می‌شود: ۱. کشف آیت‌های پرتکرار، ۲. کشف قوانین انجمنی در بین آیت‌های پرتکرار.^[۱۹]

در ابتدا با یک دور بررسی پایگاه داده، مجموعه اقلام بزرگ تک‌عضوی که تعداد تکرار آن‌ها در پایگاه داده از کم‌ترین حد تعیین شده‌ی پشتیبان بیشتر باشد، استخراج خواهد شد. سپس با استفاده از این مجموعه‌ی اقلام بزرگ تک‌عضوی، مجموعه‌ی اقلام دو‌عضوی تعیین می‌شود و برای کامل شدن فرایند جست‌وجو یک بار دیگر مجموعه اقلام دو‌عضوی در کل پایگاه داده مورد بررسی قرار می‌گیرد و مجموعه‌ی اقلام دو‌عضوی نیز به‌طور کامل استخراج می‌شوند. به همین ترتیب مجموعه‌ی اقلام سه‌عضوی و چهار‌عضوی... تعیین، و از میان آن‌ها قوانین استخراج می‌شود.

از آنجا که هدف الگوریتم اپریوری تولید مجموعه اقلام بزرگ است، در این الگوریتم پس از ایجاد مجموعه اقلام K عضو از مجموعه‌ی بزرگ K-۱ عضو

قبل از خواندن پایگاه داده برای تشخیص پوشش مجموعه اقلام K عضو، ابتدا باید برای هر مجموعه بررسی شود که آیا زیرمجموعه K-۱ عضویش بزرگ است یا نه. اگر حتی یک زیرمجموعه‌ی K-۱ عضویش هم بزرگ نباشد، آن مجموعه‌ی K عضو نمی‌تواند بزرگ باشد و باید حذف شود.

۳.۲. شبکه‌ی پیچیده

تجزیه و تحلیل‌های شبکه‌محور، شیوه‌ی مناسب برای هدایت و تحلیل داده‌ها و گروه‌های داده‌ی است که با نام شبکه‌های پیچیده مورد تجزیه و تحلیل قرار می‌گیرند. شبکه مجموعه‌ی از اشیاء است که به آن‌ها گره گفته می‌شود و بین این گره‌ها ارتباط‌هایی برقرار است که به آن‌ها یال می‌گویند. شبکه‌ی اینترنت، شبکه‌های اجتماعی، شبکه‌ی روابط کسب و کار بین سازمان‌ها، شبکه‌ی ارتباطات تلفنی، شبکه‌های عصبی... نمونه‌هایی از این شبکه‌ها هستند. شبکه‌ی که شامل یک سری یال متصل به گره‌ها باشد نمونه‌ی ساده‌ی از شبکه به‌شمار می‌رود. این شبکه‌ها براساس روش‌های مختلف تبدیل به شبکه‌ی پیچیده می‌شوند. انواع مختلفی از شبکه‌های پیچیده وجود دارد: نوع اول آن همان شبکه‌های ساده است، نوع دوم شبکه‌های با یال و گره‌های گسسته است، نوع سوم شبکه‌های با یال و گره‌های وزن‌دار، و نوع چهارم شبکه‌های جهت‌داری است که هر یال آن یک جهت دارد.^[۱۳] شبکه‌ی پیچیده از اجزای مختلفی تشکیل می‌شود:

گره: نمایان‌گر اشیاء و چیزهای مختلف در شبکه است؛

یال: هر یال یک پیوند است که دو گره مجزا را به یکدیگر وصل می‌کند؛

یال‌های جهت‌دار و غیرجهت‌دار: دو نوع اتصال عمده؛ یال‌های جهت‌دار مبدأ و مقصد معینی دارند که در شبکه به شکل یک پیکان نمایش داده می‌شود، اما یال‌های غیرجهت‌دار یک ارتباط دوسویه بین دو گره برقرار می‌کنند؛

درجه: تعداد یال‌های متصل شده به گره درجه نام دارد. شبکه‌های با یال‌های جهت‌دار دارای دو درجه‌ی ورودی و خروجی هستند؛

مؤلفه‌ها: بخش‌هایی از شبکه‌اند که به یکدیگر متصل نیستند؛

فاصله‌ی ژئودسیک: کوتاه‌ترین مسیر بین دو رأس در شبکه؛

قطر شبکه: بزرگ‌ترین فاصله‌ی ژئودسیک شبکه یا فاصله‌ی میان دو گره در شبکه که بیشترین فاصله را از هم دارند و با عددی از یک تا بینهایت نشان داده می‌شود. برای انجام تجزیه و تحلیل‌ها در شبکه از روش‌ها و معیارهای مختلفی استفاده می‌شود. مهم‌ترین این معیارها عبارت‌اند از:

-- خوشه‌بندی که شبکه را به گروه‌های خوشه‌ی مختلف موجود در شبکه تقسیم می‌کند؛

-- مرکزیت و آن درجه‌ی است که در شبکه‌های پیچیده‌ی جهت‌دار و غیرجهت‌دار به صورت مجزا بررسی می‌شود و براساس تعداد پیوندهای هر گره، گره مرکزی در شبکه معین می‌شود؛

-- مرکزیت نزدیک و آن عبارت است از فاصله‌ی میانگین میان یک گره و دیگر گره‌های شبکه؛

-- مرکزیت میانی و آن مبتنی بر این فرض است که هر یال می‌تواند با قرار گرفتن بر یک تنگ‌راه ارتباطی اقتدار کسب کند. این معیار عددی بین صفر و ۱ است که در حالت صفر با در نظر داشتن گره هیچ اتفاقی در شبکه نمی‌افتد و همه‌ی گره‌ها به یکدیگر متصل باقی می‌مانند.^[۱۳]

در تحقیقات انجام شده با ورود و تشکیل شبکه در حوزه‌های مختلف از جمله بازاریابی و فروش، شبکه‌های قاچاق، شبکه‌های روابط بین سازمانی... این رویکرد

ساختاری را در این حوزه‌ها استفاده کرده‌اند و نتایج قابل توجهی در این حوزه‌ها به دست آورده‌اند. [۱۵، ۱۴]

۴.۲. محرومیت‌زدایی از مناطق محروم

طی سال‌های متمادی، کشورهای مختلف جهان توسعه‌ی ملی کشور خود را در ارتقای درآمد سرانه و دیگر شاخص‌های توسعه‌ی اقتصادی می‌دیدند. کشورهای مختلفی این رویکرد را در اداره‌ی کشور خود داشتند که پس از آنک مدتی متوجه شدند رشد و توسعه‌ی اقتصاد کشور لزوماً نمی‌تواند به رشد و توسعه‌ی انسانی منجر شود. [۱۶] یکی از مهم‌ترین برنامه‌های کشورها رشد شاخص‌های توسعه‌ی انسانی و رفع محرومیت از مناطق کم‌توسعه‌یافته است. [۱۷] به‌لحاظ تاریخی، بررسی علمی مسئله‌ی فقر و محرومیت پس از انقلاب صنعتی و شروع به کار کارخانجات و شکل‌گیری طبقه‌ی جدید کارگر و برقراری رابطه‌ی کارفرما - کارگر در کشورهای صنعتی و نیز با گسترش پژوهش اجتماعی مورد توجه قرار گرفت. [۱۸] در ایران نیز با ورود صنایع جدید و شکل‌گیری طبقات کارگری و حاشیه‌نشینی در شهرها، قسمت‌هایی از جامعه به‌عنوان مناطق محروم شناسایی شد. توجه به این مناطق و بررسی علمی برای رفع محرومیت این مناطق، می‌تواند در رشد و توسعه‌ی کشورها بسیار حائز اهمیت باشد.

۵.۲. کاربرد فنون داده‌کاوی در محرومیت‌زدایی از مناطق محروم

با توجه به ماهیت کاربردی فنون داده‌کاوی که دانش‌ها و روابط پنهان میان پایگاه داده‌ها را کشف می‌کند، با به کار بردن این فنون در داده‌های جمع‌آوری شده و آماده‌سازی شده از نحوه‌ی محرومیت مناطق مختلف می‌توان دانش‌ها و اطلاعات مفید و قابل توجهی را استخراج کرد و در اختیار تصمیم‌گیرندگان در این حوزه قرار داد. از این رو از فنون پیش‌بینانه و توصیفی داده‌کاوی، به‌خصوص فنون خوشه‌بندی، برای تحلیل و بررسی خوشه‌ی کشورهای بر روی داده‌های محرومیت‌های مناطق استفاده شده است. در جدول ۱ نمونه‌ی از تحقیقات انجام گرفته و کاربرد فنون داده‌کاوی در حوزه‌ی محرومیت‌زدایی از مناطق محروم ارائه شده است.

جدول ۱. خلاصه‌ی از کاربردهای داده‌کاوی در حوزه تجزیه و تحلیل محرومیت.

مرجع	فنون داده‌کاوی	الگوریتم داده‌کاوی	خلاصه کاربرد
[۵]	خوشه‌بندی	کا - میانگین	تمرکز بر شاخص سلامت و پراکندگی جغرافیایی در محرومیت و تجزیه و تحلیل خوشه‌ی آنها
[۱۷]	خوشه‌بندی	نقشه‌های خودسازمانده	شاخص سلامت و برنامه‌ریزی شهری و چگونگی خوشه‌های تشکیل‌شده‌ی این شاخص بر روی نقشه‌های سیستم اطلاعات جغرافیایی
[۹]	قوانین انجمنی و کشف توالی	کشف الگوهای دنباله‌ی	روابط بین محرومیت منطقه‌ی و ریسک سرطان
[۶]	خوشه‌بندی	کا - میانگین	کشف مهم‌ترین شاخص‌ها براساس جهت‌گیری خوشه‌ی و تأثیرگذاری آنها
[۷]	رگرسیون	مدل پیش‌بینانه رگرسیون	پیش‌بینی همبستگی بین شاخص اصلی محرومیت با شاخص‌های جنسیت، درآمد، سلامت و نرخ مراجعه به بیمارستان
[۱۶]	خوشه‌بندی و طبقه‌بندی	کا - میانگین و شبکه عصبی	تجزیه و تحلیل و ایجاد مدل براساس شاخص‌های توسعه‌ی انسانی، باسوادی، درآمد سالانه، امید به زندگی و شاخص محرومیت در کشورهای اسلامی
[۸]	خوشه‌بندی	کا - میانگین	ارائه‌ی استراتژی برای رفع محرومیت مناطق کشور سوییس براساس جهت‌گیری‌های خوشه‌ی شاخص‌های چندگانه محرومیت این کشور

۳. بیان مسئله

در تمامی کشورها با توجه به نرخ و میزان محرومیت برخی از نواحی، مناطقی وجود دارد که به‌لحاظ امکانات و تسهیلات در دسترس و براساس شاخص‌های محرومیت و توسعه‌ی انسانی، در ردیف مناطق محروم و کم‌توسعه‌یافته قرار می‌گیرند. در گذشته، در جوامع صنعتی این تفکر وجود داشت که تنها نیاز کشورها دست‌یابی به درآمد بالا و توسعه‌ی اقتصادی است، اگرچه با وجود تمرکز بر توسعه‌ی اقتصادی نتوانستند نیازهای اقشار مختلف جامعه‌ی خود را به‌طور یکسان برطرف کنند. به همین دلیل در آن سال‌ها این نتیجه حاصل شد که توسعه‌ی اقتصادی لزوماً به توسعه‌ی انسانی منجر نمی‌شود. [۱۶] اما با هدف‌مند شدن و تمرکز بر متغیرهای توسعه‌ی انسانی براساس دانش‌های پنهان و تخصیص بودجه‌ی هدف‌مند براساس این نتایج، مزیت توسعه‌ی اقتصادی این کشورها نوعاً می‌تواند به ارتقاء و توسعه‌ی انسانی آنها کمک کند.

بحث محرومیت‌زدایی از مناطق محروم موضوع مهمی است که در بیشتر مجامع علمی مورد توجه قرار دارد. در تمام کشورهای ارتقای شاخص‌های توسعه‌ی انسانی و کاهش شاخص‌های محرومیت و روش‌های مختلف دسترسی به آنها، در دستورکار قرار دارد. [۲] داده‌کاوی و کشف دانش و استفاده از فنون مختلف آن روشی است که در تحقق این موضوع مهم نقش به‌سزایی دارد. داده‌کاوی با کاوش و جست و جو در میان حجم انبوه داده‌ها، به کشف روابط پنهان در میان این داده‌ها می‌پردازد و آن را برای به‌کارگیری در استراتژی‌ها و راهبردهای مختلف آماده می‌کند. نکته‌ی قابل توجه این است که داده‌کاوی، روابط باهمایی که در بین شاخص‌های محرومیت در مناطق مختلف وجود دارد را شناسایی و آنها را برای اقدام هرچه بهتر و مفیدتر در جامعه معین خواهد کرد. برای این کار به یک سری داده‌ها و شاخص‌های مبین محرومیت مناطق مختلف نیازمندیم تا بتوانیم به‌کمک آنها و فرایند داده‌کاوی راهکارهایی برای بهبود وضعیت مناطق مختلف محروم ارائه دهیم. یکی از مهم‌ترین و پرکاربردترین این شاخص‌ها، شاخص توسعه‌ی انسانی (HDI) [۴] است که میزان توسعه‌یافتگی کشورهای مختلف در جهان را مشخص می‌کند. [۱۹] این شاخص به‌دلیل بزرگ بودن در محدوده‌ی جغرافیایی مورد بررسی، برای سازوکار تحقیق ما مناسب نیست. شاخص

براساس این متدولوژی در گام اول به شناخت جایگاه سیستم و اهداف و چشم‌اندازهای آن خواهیم پرداخت. در گام شناخت سیستم به نوعی مطالعات امکان‌سنجی و شناخت منابع پروژه داده‌کاوی مورد تجزیه و تحلیل قرار خواهد گرفت. سپس با استفاده از این دانش‌ها تعریفی از مسئله‌ی داده‌کاوی و طرح مقدماتی ارائه خواهد شد.

گام دوم این متدولوژی در ابتدا با جمع‌آوری داده‌های اولیه شروع می‌شود و سپس با انجام کارهایی به منظور آشنایی بیشتر مسئول پروژه با نوع داده‌ها و ویژگی‌های آن ادامه می‌یابد. در این گام، اطلاع از کیفیت و اعتبار داده‌ها منوط به وجود شناخت جامع و دقیق از آن‌هاست؛ برای محقق شدن این امر بایستی جمع‌آوری داده‌های اولیه، توصیف داده‌ها، بازرسی و بررسی داده‌ها و در نهایت اعتبارسنجی کیفیت داده‌ها به‌طور کامل انجام شود.

«آماده‌سازی داده‌ها» مرحله‌ی سوم از متدولوژی داده‌کاوی است. این گام همه‌ی فعالیت‌هایی را که نیازمند ساخت داده‌های نهایی از داده‌های خام است تحت پوشش قرار می‌دهد. فرایندهای مختلف این مرحله عبارت‌اند از: انتخاب داده‌ها، پاک‌سازی داده‌ها، آماده‌سازی داده‌ها برای اجرای الگوریتم‌ها و روش‌های داده‌کاوی، تجمیع داده‌ها و در نهایت قالب‌بندی داده‌ها.

در مرحله‌ی چهارم فنون مختلف موجود در داده‌کاوی براساس تعیین مسئله و اهداف تحقیق، انتخاب و به کار گرفته خواهد شد. در این مرحله پارامترهای مدل‌های مختلف تعیین و به کار گرفته خواهد شد. کیفیت خروجی فرایند داده‌کاوی بستگی زیادی به انتخاب روش و الگوریتم مناسب برای داده‌ها در این مرحله دارد.

مرحله پنجم به تفسیر الگوها و قوانین به دست آمده اختصاص دارد. برای انجام این ارزیابی‌ها از نظرات کارشناسان خبره در این زمینه استفاده می‌شود؛ همچنین مشاهدات نتایج و روابط منطقی که در مدل وجود دارد توسط خود فرد محقق انجام می‌شود. قبل از ارائه‌ی نهایی مدل، این مرحله باید حتماً روی مدل‌های انتخابی و نتایج آن صورت گیرد و کارایی آن براساس اهداف مد نظر ما تأیید شود. در مرحله‌ی پایانی تصمیم درمورد استفاده از داده‌کاوی باید به نتیجه برسد. اگر بعد از کشف دانش در مراحل قبلی به دانش‌هایی مفید و کامل برسیم، این دانش‌ها نیازمند سازمان‌دهی، واری و بازنگری خواهند بود که این اقدامات در گام ششم انجام خواهد شد.^[۱۰]

علاوه بر استفاده از روش کریسپ داده‌کاوی برای اجرای این تحقیق، از فنون شبکه‌های پیچیده و تجزیه و تحلیل آن‌ها برای استخراج نتایج بیشتر و هرچه بهتر براساس الگوریتم پیشنهادی استفاده خواهد شد. در جدول ۲ خلاصه‌ی این مراحل لازم برای اجرای این تحقیق بر روی پایگاه داده‌ی شاخص‌های محرومیت تشریح شده است.

۵. پیاده‌سازی چارچوب پیشنهادی برای اجرای تحقیق

۱.۵. شناخت سیستم

در این مرحله مفهوم محرومیت‌زدایی و شناخت کلی سیستم را مورد بررسی قرار داده‌ایم. شناسایی مناطق محروم و همچنین شناسایی نوع محرومیت این مناطق یکی از موضوعات مهمی است که در برنامه‌های مهم کشورها جایگاه ویژه‌ی دارد. سازمان جوامع و شهرداری‌های انگلستان یکی از سازمان‌های پیشرو در این زمینه است. آن‌ها با ارائه‌ی شاخص‌های چندگانه‌ی محرومیت و انجام تجزیه و تحلیل‌های آماری روی آن به ارزیابی و سنجش کیفیت زندگی در مناطق مختلف، و مقایسه‌ی

مهم و پرکاربرد دیگری که در این زمینه مورد استفاده قرار می‌گیرد شاخص‌های چندگانه محرومیت (IMD)^۵ است که معمولاً برای سنجش محرومیت مناطق کوچک به کار می‌رود. این شاخص از سوی سازمان جوامع و شهرداری‌های انگلستان ارائه شده و در واقع کامل‌ترین و دقیق‌ترین شاخص در زمینه‌ی سنجش محرومیت در مناطق کشورهاست.^[۲۰]

چنان‌که ذکر شد در کشورهای مختلف مکرراً اقداماتی در راستای محرومیت‌زدایی از مناطق کم‌تر توسعه‌یافته و محروم انجام می‌گیرد. این اقدامات به‌صورت تجربی و براساس گزارشاتی که طبق شاخص‌های چندگانه‌ی محرومیت ارائه می‌شود، انجام می‌گیرد. پس از مشخص شدن میزان محرومیت هر منطقه، به آن دسته از مناطقی که با هر شاخصی «محروم» به شمار می‌رود یک ردیف بودجه‌ی اختصاص می‌یابد تا آن منطقه از نظر آن شاخص تقویت شود. این مقاله به نوعی بیان‌گر یک رویکرد خاص در آماده‌سازی داده‌های شاخص‌های چندگانه‌ی محرومیت و تحلیل آن‌ها به‌منظور ارتقای بهره‌وری در اختصاص بودجه‌های مختلف در راستای محرومیت‌زدایی مناطق است. هدف اصلی از اجرای این تحقیق بهبود در وضعیت ارتقای شاخص‌های محرومیت و محرومیت‌زدایی از مناطق محروم است. به‌طور کلی اهداف مد نظر عبارت‌اند از:

۱. کشف و استخراج روابط پنهان بین شاخص‌های محرومیت و کشف متغیرهای مهم در محرومیت یک منطقه؛
۲. کشف و استخراج مهم‌ترین متغیرها در محرومیت‌زدایی جزئی و کلی در مناطق مختلف کشورها براساس شبکه‌ی پیچیده؛
۳. ارائه‌ی چارچوب و پیشنهاداتی براساس روابط منطقی موجود برای رسیدن به سطح رفاه اجتماعی بالاتر؛
۴. استفاده از نتایج به دست آمده در راستای کاهش سطح محرومیت مناطق مختلف. در این تحقیق سعی می‌شود تا با به‌کارگیری روش‌های ابتکاری در آماده‌سازی داده‌ها و پیاده‌سازی قوانین انجمنی در شاخص‌های چندگانه‌ی محرومیت، الگوهای پنهان موجود در بین این شاخص‌ها را کشف و از آن‌ها برای تحقق اهداف مورد نظر در محرومیت‌زدایی هدف‌مند و هرچه بهتر در کشورها بهره برد.

۴. متدولوژی اجرای تحقیق

متدولوژی و چارچوب کلی اجرای این تحقیق براساس متدولوژی استاندارد داده‌کاوی کریسپ^۶ است. این روش به‌علت برخورداری از خاصیت‌های جامعیت و مانعیت و داشتن رویه‌ی کامل برای انجام این نوع تحقیقات از محبوبیت خاصی برخوردار شده است.^[۱۱] در این تحقیق نیز با توجه به مسئله‌ی موجود و روش تحقیق مورد بررسی، برای رسیدن به اهداف مورد نظر و پاسخ‌گویی به سؤالات، از این روش استفاده شده است. به‌طور کلی این روش دارای شش گام اصلی است:

۱. شناخت سیستم؛
۲. شناخت داده‌ها؛
۳. آماده‌سازی داده‌ها؛
۴. مدل‌سازی؛
۵. ارزیابی؛
۶. توسعه.^[۱۱]

جدول ۲. چارچوب پیشنهادی برای اجرای تحقیق.

شناخت سیستم و آشنایی با حوزه‌ی محرومیت‌زدایی و شاخص‌های محرومیت و چگونگی برخورد با آن	گام اول شناخت سیستم
چگونگی دسترسی به داده‌های شاخص‌های چندگانه محرومیت و شناخت داده‌ها و ویژگی‌های این پایگاه داده	گام دوم شناخت داده‌ها
نرمال‌سازی و ارائه‌ی الگوریتم ابتکاری برای گسسته‌سازی داده‌های شاخص‌های چندگانه محرومیت و آماده کردن داده‌ها	گام سوم آماده‌سازی داده‌ها
اجرای الگوریتم اperiوری روی داده‌ها و استخراج قوانین باهمایی موجود در پایگاه داده و تجزیه و تحلیل آن‌ها، ایجاد شبکه قوانین براساس یک الگوریتم ابتکاری و کشف تأثیرگذارترین و تأثیرپذیرترین شاخص‌ها بر روی مناطق و تحلیل آن‌ها	گام چهارم مدل‌سازی
بررسی و ارزیابی نتایج به دست آمده توسط افراد خبره در این حوزه و بررسی مجدد فرایند تحقیق	گام پنجم ارزیابی
توسعه و سازمان‌دهی نتایج و تهیه‌ی طرح پروژه	گام ششم توسعه

روش پیشنهادی و کمیته استفاده می‌کنیم:

$$S' = \frac{(S - S_{\min})}{(S_{\max} - S_{\min})}$$

با اعمال این رابطه روی تمامی شاخص‌ها، داده‌ها در بازه بین صفر و ۱ قرار می‌گیرند.

۱.۳.۵. گسسته‌سازی داده‌ها

پس از یکسان‌سازی مقیاس داده‌ها توسط روش‌های نرمال‌سازی، داده‌ها را باید آماده‌ی اجرای الگوریتم‌های بخش گسسته‌سازی کرد. پس از انجام گسسته‌سازی بر روی یک فیلد، فیلد جدیدی با مقادیر جدید و با عرض ثابت یا پویا به مجموعه پایگاه داده اضافه می‌شود. به‌عنوان مثال اگر داده‌های شاخص درآمد مقادیر عددی بین صفر و ۱ داشته باشد، با ایجاد یک فیلد جدید -- شامل گروه‌های درآمدی مختلف با عرض یکسان یا عرض پویا -- گسسته‌سازی می‌شود.^[۲۱]

برای گسسته‌سازی داده‌های نسبتی نرمال شده دو رویکرد اصلی و کلی وجود دارد. رویکرد اول مربوط به روش‌های با ناظر^۲ است. در این روش‌ها با استفاده از یک فیلد طبقه‌بندی به‌عنوان فیلد ناظر، الگوریتم مورد نظر اجرا می‌شود. به جهت اینکه تمام فیلدهای این پایگاه داده از نوع داده‌های نسبتی و بین صفر و ۱ هستند لذا نمی‌توان هیچ‌کدام از آن‌ها را به‌عنوان فیلد ناظر انتخاب کرد و باید از روش‌های دیگر استفاده کرد.

رویکرد بعدی که در گسسته‌سازی داده‌های نسبتی وجود دارد روش‌های بدون ناظر^۳ است. این روش‌ها احتیاجی به فیلد ناظر ندارند. در این رویکرد دو الگوریتم وجود دارد که اولین الگوریتم روش گسسته‌سازی است که داده‌ها را براساس فراوانی شان دسته‌بندی می‌کند. روش دوم الگوریتم گسسته‌سازی است که داده‌ها را در بازه‌هایی با طول یکسان دسته‌بندی می‌کند.^[۲۲] این روش به نوعی برای داده‌های پایگاه داده ما مناسب است اما نکته‌ی جالب توجه اینجاست که اهمیت مناطقی که در شاخص‌های مختلف از محرومیت بیشتری برخوردارند، برای ما بیشتر است. از سوی دیگر در قوانین انجمنی ما به دنبال کشف روابط پنهان بین مناطق محروم‌تر هستیم، لذا باید الگوریتمی طراحی شود که بازه‌های ابتدایی نشان‌گر محرومیت بیشتر نسبت به آن شاخص، رنج و طول بازه بیشتری داشته باشد و رفته رفته کم شده و به حالت عادی روش گسسته‌سازی که فاصله‌ی ثابت دارد برسد.

روش ابتکاری گسسته‌سازی با فاصله پویا. در این روش گسسته‌سازی، از طریق مراحل زیر نقاط انفصال برای گسسته‌سازی داده‌ها معین می‌شود؛ G_i ها بیان‌گر این نقاط انفصال اند:

الف) داده‌ها را به‌صورت صعودی مرتب کنید.

آن‌ها را با یکدیگر می‌پردازند. در واقع آن‌ها با انجام این تحلیل‌ها به بررسی خوشه‌ی و تحلیل چگونگی وضعیت هر شاخص در هر منطقه می‌پردازند. ما از این داده‌ها به‌عنوان داده‌های دست دوم برای رسیدن به اهداف و پاسخ‌گویی به سؤالات این تحقیق استفاده کرده‌ایم.

به‌طور کلی حوزه‌ی مورد بررسی این تحقیق محرومیت‌زدایی بهینه و هدف‌مند از مناطق مختلف کشور است.

۲.۵. شناخت داده‌ها

داده‌های مورد استفاده در این تحقیق از پایگاه اطلاع‌رسانی سازمان جوامع و شهرداری‌های انگلستان برداشت شده است. این پایگاه داده ۹ شاخص کلی را شامل می‌شود: درآمد، اشتغال، سلامت و معلولیت، آموزش مهارت‌ها، دسترسی به مسکن، دسترسی به خدمات، میزان ارتکاب جرم، شرایط محیط‌زیست و زندگی. هر یک از این شاخص‌های کلی چند زیرشاخص از جنس خودشان دارد که در نهایت پایگاه داده‌ی ما ۲۵ شاخص را تشکیل می‌دهند. همچنین این سازمان، کشور انگلستان را به ۳۲۴۸۲ منطقه تقسیم کرده و برای هر منطقه براساس سنجه‌های از پیش تعیین شده یک امتیاز به هر کدام از شاخص‌ها اختصاص می‌دهد. در واقع ۳۲۴۸۲ منطقه به‌عنوان رکوردهای این پایگاه داده و ۲۵ شاخص به‌عنوان فیلدهای این پایگاه داده شناخته می‌شوند.

۳.۵. پیش‌پردازش و آماده‌سازی داده‌ها

این مرحله‌ی داده‌کاوی مهم‌ترین و زمان‌برترین مرحله‌ی پروژه‌های داده‌کاوی است. یکی از مراحل مهم پیش‌پردازش داده‌ها انتخاب فیلدها و مشخصه‌های مناسب است. در این زمینه با در نظر گرفتن تحقیقات انجام شده و ادبیات موجود و همچنین نظر خبرگان و تحلیل‌های کمی انجام شده در خصوص ارتباط و همبستگی بین مشخصه‌ها، تمامی ۲۵ مشخصه‌ی در دسترس که هر کدام یک نوع محرومیت معین را شامل می‌شود، برای استخراج قوانین به کار گرفته شده‌اند.

اما در بررسی نوع داده‌ها، در این پایگاه داده امتیازهایی به هر منطقه و شاخص براساس سنجه‌ها داده شده است. هر کدام از این شاخص‌ها در بازه عددی متفاوتی قرار دارد و بنابراین برای یکسان‌سازی و تغییر مقیاس هر کدام از شاخص‌ها از روش نرمال‌سازی استفاده کرده‌ایم.

برای نرمال‌سازی داده‌ها و حفظ رابطه بین مقادیر داده‌های اصلی و تبدیل داده‌ها بین عددی صفر و ۱، با توجه به مثبت یا منفی بودن ماهیت متغیر، از

جدول ۳. تعداد دسته‌های هریک از شاخص‌های کلی.

تعداد دسته‌ها	شاخص کلی
۳۵	شاخص چندگانه‌ی محرومیت
۶۸	درآمد
۴۹	اشتغال
۶۷	سلامت
۵۱	آموزش
۴۷	ارتکاب جرم و جنایت
۳۴	محدودیت‌های در دسترس زندگی
۱۱۵	محیط زیست و زندگی
۶	جمعیت

۱.۴.۵. تفسیر نتایج استخراج شده

گزیده‌یی از نتایج حاصل از اجرای الگوریتم اپریوری بر روی داده‌ها در جدول ۴ آمده است. هرکدام از این قوانین به دست آمده یک تفسیر دارند؛ در این بخش با توجه به متدولوژی در نظر گرفته شده برای اجرای این تحقیق، گام‌های ارزیابی و توسعه‌ی مدل براساس نتایج به دست آمده انجام شده و نتایج معنادار به همراه تفسیرهای آن‌ها در زیر آمده است:

به‌عنوان مثال قانون شماره ۱ به این موضوع اشاره دارد که آن دسته از مناطقی که از نظر شاخص نرخ بیماری و معلولیت -- که یکی از مهم‌ترین شاخص‌های سلامت است -- از وضعیت خوبی برخوردار باشند، از نظر شرایط اقلیمی زندگی در محیط‌های مناسب باشند، و نیز از نظر آموزش و مهارت‌های فنی و تحصیلات در رتبه‌ی مناسبی قرار بگیرند، با اطمینان ۱۰۰٪ از نظر شاخص اصلی چندگانه‌ی محرومیت در رتبه‌ی ۱ (یعنی در رده‌ی مرفه‌ترین مناطق از منظر این شاخص‌ها) قرار می‌گیرند. این قانون به نوعی بیان‌گر این نکته است که برای بالا بردن سطح رفاه مناطقی که از نظر شاخص IMD در رتبه‌ی مناسبی نیستند، می‌توان بر سه شاخص تحصیلات، شرایط اقلیمی محیط زندگی، و بیماری و معلولیت متمرکز شد و بیش از پیش به ارتقای سطح رفاه مناطق کمک کرد.

قانون شماره ۲ نیز بیان‌گر این موضوع است که آن دسته از مناطقی که از نظر شاخص سلامت و برخورداری از نعمت جسم سالم دچار محرومیت‌اند، و در حد معمول و میانگین این شاخص قرار دارند اما از نظر شاخص درآمد در رتبه‌ی خوبی قرار دارند، با اطمینان ۱۰۰٪ از نظر شاخص IMD در رتبه‌ی ۱ و جزء مناطق با وضعیت مناسب قرار می‌گیرند. این قانون بیان می‌دارد که مناطق مرفه که افرادی با وضعیت اقتصادی خوب و مناسب در آن زندگی می‌کنند اگر دچار بیماری شوند یا از لحاظ سلامتی دچار مشکل شوند با داشتن مزیت وضعیت اقتصادی خوب می‌توانند به درمان و بهبود وضعیت سلامتی خود بپردازند. در واقع این شاخص درآمد خوب، می‌تواند پوشش مناسبی برای شاخص سلامت باشد و افراد را در وضعیت IMD بالا نگه دارد. نکته‌ی دیگری که می‌توان در این‌جا به آن اشاره کرد این است که اگر دولت بخواهد اوضاع شاخص سلامت و به تبع آن شاخص IMD را بهبود بخشد، افزایش سطح درآمد مناطق کم‌درآمد یکی از آن راه‌ها است.

به‌عنوان مثالی دیگر قانون شماره ۴ بیان می‌دارد که آن دسته از مناطقی که از نظر شاخص سلامت وضعیت خوبی ندارند و همچنین از نظر شاخص فاصله تا مرکز بهداشت، پزشک عمومی و بیمارستان از وضعیت مناسبی برخوردار نیستند،

ب) بیشینه و کمینه‌ی داده‌ها را پیدا کنید.

ج) مقدار k (تعداد مقادیر گسسته‌سازی) را توسط افراد خیره مشخص کنید؛

د) مقدار $\partial = \frac{(X_{\max} - X_{\min})}{k}$ را محاسبه کنید.

ه) حدود گسسته‌سازی عبارت خواهد بود از:

$$G_i = X_{\min} + i * \partial \quad i = 1, 2, 3, \dots, k - 1$$

براین اساس داده‌ها با طول بازه یکسان و برابر گسسته‌سازی می‌شوند. اما در روش ابتکاری ما سه متغیر دیگر به نام‌های D و F و R معرفی می‌کنیم؛ براساس الگوریتم زیر متغیر R بیان‌گر نقاط انفصال با بازه‌های متغیر است:

$$D_i = (G_{i+1} - G_i) + F_i$$

که:

$$F_1 = \frac{\partial}{2},$$

$$F_i = \frac{(F_{i-1})}{2}, \quad i = 2, \dots, k - 1,$$

$$R_i = (D_{i-1}) + D_i$$

که:

$$D_0 = X_{\min}$$

به این ترتیب مقدار بازه‌ها و نقاط انفصال تغییر خواهد کرد و به شرایط مورد نظر ما خواهد رسید. R_i بیان‌گر نقاط انفصال داده‌ها براساس الگوریتم ابتکاری خواهد بود. بنابراین بازه‌های ابتدایی که گسسته شده‌اند بزرگ‌ترند و رفته رفته کم می‌شوند و به مقدار ثابت روش گسسته‌سازی می‌رسند.

با به‌کار بردن این الگوریتم ابتکاری در گسسته‌سازی داده‌های شاخص‌های چندگانه محرومیت، تمامی این شاخص‌ها در قالب متدولوژی این الگوریتم گسسته‌سازی و دسته‌بندی می‌شوند. مقدار K یا همان تعداد دسته‌ها در هر گروه از شاخص‌های کلی براساس جهت‌گیری‌های خوشه‌یی که هرکدام از شاخص‌های کلی در بین خود دارند تعیین و به کار گرفته شده‌اند. این مقادیر طبق این جهت‌گیری‌ها در جدول ۳ آمده است.

به این ترتیب داده‌های این پایگاه داده گسسته‌سازی می‌شود. به‌عنوان مثال در شاخص‌هایی که زیرمجموعه‌ی شاخص کلی ارتکاب جرم و جنایت باشند براساس الگوریتم مطروحه، به هر منطقه یک رتبه بین ۱ تا ۴۷ تعلق می‌گیرد؛ رتبه‌های اولیه که بازه بزرگ‌تری دارند نشان از وضعیت بحرانی به لحاظ وقوع جرم و جنایت توسط میانگین اقشار آن منطقه دارد. دیگر شاخص‌ها نیز به این ترتیب گسسته‌سازی، و برای اجرای الگوریتم‌های مدل‌سازی آماده می‌شوند.

۴.۵. مدل‌سازی داده‌ها و دانش‌های حاصل از مدل‌سازی

در بخش قبل داده‌ها با روش‌های نرمال‌سازی و گسسته‌سازی که روی آن‌ها انجام شد، آماده‌ی اجرای الگوریتم‌های مدل‌سازی شدند. بنابراین با اجرای الگوریتم اپریوری از روش قوانین انجمنی، روابط پنهان و وابستگی‌های بین شاخص‌های محرومیت به‌منظور استفاده‌ی بهینه در تصمیم‌گیری برای محرومیت‌زدایی از مناطق محروم شناسایی می‌شود. برخی از نتایج معنادار در جدول ۴ ارائه شده است.

جدول ۴. نمونه‌ی از قوانین انجمنی استخراج شده.

شماره	تالی	مقدم	اطمینان	پشتیبان	صعود
۱	نمره شاخص چندگانه محرومیت = ۱	بیماری و معلولیت = ۷ شرایط محیط زندگی = ۱ تحصیلات و مهارت‌های فنی = ۱	۱۰۰	۲,۱۲۱۱	۳,۰۴
۲	نمره شاخص چندگانه محرومیت = ۱	درآمد = ۲۸ سلامت و معلولیت = ۳۲	۱۰۰	۱,۱۱۱۴	۳,۰۳
۳	نمره شاخص چندگانه محرومیت = ۱	شرایط محیط زندگی = ۱ تحصیلات و مهارت‌های فنی = ۱ جمعیت کل = ۳	۹۶,۱۷	۶,۷۳۹۱	۲,۹۲
۴	محرومیت‌های خانه و سرویس‌های در دسترس آن = ۱	فاصله تا پزشک عمومی = ۳ سلامت و معلولیت = ۳۲	۹۱,۹۶	۱,۹۰۲۵	۱,۸۷
۵	نمره شاخص چندگانه محرومیت = ۲	درآمد = ۴۹	۷۶,۰۴	۲,۱۲۱۱	۳,۳۶
۶	استطاعت مالی = ۱	آلودگی هوا = ۱۶ محرومیت‌های خانه و سرویس‌های در دسترس آن = ۱	۷۵	۰,۱۲۲۴	۱۹,۱
۷	بیماری و معلولیت = ۹	اشتغال = ۳۴ تحصیلات و مهارت‌های فنی = ۱ جمعیت کل = ۳	۴۶,۱۷	۱,۰۴۰۵	۴,۵۸
۸	نمره شاخص چندگانه محرومیت = ۳	درآمد = ۵۴ جمعیت کل = ۳ درآمد = ۶۵	۴۳,۸۷	۱,۵۱۱۶	۲,۶۵
۹	محرومیت از درآمد تأثیرگذار بر کودکان = ۶۶	محرومیت‌های خانه و سرویس‌های در دسترس آن = ۱ جمعیت کل = ۳ جمعیت خردسال بین ۰ تا ۱۵ سال = ۱	۵۷,۲۲	۰,۲۹۲۴	۲۵,۹
۱۰	ترک تحصیل = ۵۰	محرومیت از درآمد تأثیرگذار بر کودکان = ۶۶ اشتغال = ۴۶ شرایط محیط زندگی = ۱	۴۵,۴۳	۰,۰۶۴۶	۴۴,۸
۱۱	سال‌های بالقوه زندگی = ۳۲	عدم ورود به تحصیلات عالی = ۲۶ درآمد = ۴۹ جمعیت کل = ۳	۶۴,۷۰	۰,۰۳۳۸	۷,۵۱
۱۲	بیماری و معلولیت = ۷	اشتغال = ۳۰ عدم ورود به آموزش عالی = ۳ جمعیت جوان و خردسال = ۱	۶۳,۱۱	۰,۲۳۷۰	۶,۶۸
۱۳	آلودگی هوا = ۳۹	نیترژن دی اکسید = ۴۵ شاخص ذرات معلق = ۳۸ سولفور دی اکسید = ۹۹	۱۰۰	۰,۲۰۰۱	۴۴,۱
۱۴	محرومیت‌های جغرافیایی = ۱	فاصله تا مدرسه ابتدایی = ۸	۱۰۰	۰,۹۸۲۸	۹,۲۴
۱۵	شرایط محیط زندگی = ۱	شاخص تصادفات ترافیکی = ۲۳ نمره شاخص چندگانه محرومیت = ۱ نمره شاخص چندگانه محرومیت = ۱۳	۷۸,۰۴	۰,۱۹۷۰	۵,۵۴
۱۶	درآمد = ۶۳	اشتغال = ۴۵ بنزن در هوا = ۱۰۰ جمعیت کل = ۳	۶۴	۰,۰۴۹۲	۲۳,۱
۱۷	سلامت و معلولیت = ۳۲	شاخص مرگ و میر = ۳۳ بنزن در هوا = ۸۸	۸۸,۲۳	۰,۲۳۰۳	۰,۸۸
۱۸	اشتغال = ۴۳	درآمد = ۵۶ بیماری و معلولیت = ۲۷	۶۵	۰,۰۸۰۱	۶,۵۸

ادامه‌ی جدول ۴.

شماره	تالی	مقدم	اطمینان	پشتیبان	صعود
۱۹	جرم و جنایت = ۲۸	محرومیت از درآمد تأثیرگذار بر کودکان = ۶۵ سال‌های بالقوه زندگی = ۳۹ جمعیت کل = ۳	۸۹٫۱۸	۰٫۱۰۱۵	۰٫۸۹
۲۰	جرم و جنایت = ۲۰	محرومیت از درآمد تأثیرگذار بر افراد مسن = ۲۱ تحصیلات و مهارت‌ها = ۲ درآمد = ۱	۵۲٫۶۳	۰٫۰۳۰۷	۱۰٫۱

فراهم نشود، دچار بیماری‌های ناشی از افسردگی یا اعتیاد و غیره خواهند شد. در تفسیر قانون شماره ۱۹ نیز باید گفت که آن دسته از مناطقی که ساکنین آن‌ها کم‌جمعیت یا دارای جمعیت متوسط‌اند و همچنین از نظر شاخص درآمد تأثیرگذار بر کودکان در مضیقه‌اند، فرزندان آن‌ها دچار مشکلات‌اند و سال‌های بالقوه زندگی خود را برای کسب درآمد و کسب علم از دست داده‌اند، با اطمینان ۸۹٫۱۸٪ از نظر شاخص اعمال جرم و وقوع جرم در جایگاه نامناسبی قرار می‌گیرند. به عبارت دیگر مناطقی که این ویژگی‌ها را داشته باشند پتانسیل ارتکاب جرم در آن‌ها زیاد است.

۲.۴.۵. ایجاد و تحلیل شبکه‌ی قوانین

در این مقاله ما با رویکردی نوین از تحلیل شبکه‌های پیچیده در تفسیر هرچه بهتر نتایج داده‌کاوی استفاده خواهیم کرد. برای این کار با تشکیل شبکه قوانین انجمنی براساس الگوریتم ابتکاری طراحی شده، این شبکه طراحی و تجزیه و تحلیل خواهد شد.

به‌منظور تحلیل هرچه بهتر قوانین کشف شده از بخش قوانین انجمنی و همچنین کشف مهم‌ترین و تأثیرگذارترین و تأثیرپذیرترین شاخص‌ها، به طراحی یک الگوریتم ابتکاری در راستای ایجاد شبکه‌ی پیچیده‌ی قوانین خواهیم پرداخت. ما به دنبال کشف شاخص‌هایی هستیم که بیشترین ورودی و بیشترین خروجی را در شبکه‌ی طراحی شده دارند. لذا الگوریتم ابتکاری طراحی شده برای تشکیل شبکه‌ی قوانین و استخراج روابط میان آن عبارت است از:

- تمام قوانین با بخش مقدم و تالی را که تعداد آیتم‌های آن‌ها یک مورد است، استخراج می‌کنیم.
 - معیار اندازه‌گیری اطمینان را برای هر یک از قوانین به دست می‌آوریم و آن‌ها را به‌صورت نرمال درمی‌آوریم.
 - در شبکه‌ی ما نام آن را شبکه‌ی قوانین می‌نامیم، هرکدام از اجزای منحصر به فرد قوانین یعنی شاخص و رتبه‌ی مربوط به آن، به‌عنوان یک گره در نظر گرفته می‌شود و برای هر قانون از گره بخش مقدم یک یال به گره بخش تالی رسم می‌شود.
 - وزن این یال‌ها را برابر با معیار اطمینان که نرمال شده، در نظر می‌گیرند.
 - شبکه رسم و گره‌های مرکزی و دیگر نتایج مربوط به این شبکه استخراج خواهد شد.
- در تجزیه و تحلیل انجام شده، تعداد قوانین انجمنی کشف شده که به‌صورت تک‌جزه هستند و شامل حال الگوریتم شبکه پیچیده می‌شوند، حدود ۹۰۰۰۰ قانون است و ۸۴۴ گره منحصر به فرد نیز کشف شده است. همان‌طور که در الگوریتم آمده، از هرکدام از گره‌ها یک یال با وزن معیار اطمینان نرمال شده، به گره دیگر رسم می‌شود.

با اطمینان ۹۱٫۹۶ درصد از نظر محدودیت‌های منزل مسکونی خود و کمبودهای سرویس‌های در دسترس، از وضعیت خوبی برخوردار نخواهند بود. این قانون نیز به‌نوعی بیانگر اهمیت سلامت افراد و مراکز مرتبط با سلامت در معیار اندازه‌گیری وضعیت محدودیت‌های خانه‌ها و سرویس‌های در دسترس آن مناطق است. به عبارت دیگر این قانون بیان می‌دارد که در مناطق مختلف بیش از همه‌ی سرویس‌هایی که باید در دسترس مردم باشد، سلامت و مراکز مرتبط با آن — نظیر بیمارستان‌ها و مراکز بهداشت و پزشکی عمومی — در اولویت قرار دارد.

در نمونه‌ی دیگر، قانون شماره ۷ این‌گونه تفسیر می‌شود که مناطقی که از نظر تحصیلات و مهارت‌های فنی در وضعیت مناسبی هستند و البته در آن خانواده‌های کم‌جمعیت جای دارند اما وضع اشتغال آن‌ها یک مقدار از حد معمول و خوب فاصله دارد و تقریباً نامناسب است، با اطمینان ۴۶٫۱۷ درصد وضع شاخص بیماری و معلولیت تحت تأثیر قرار می‌گیرد و از حالت نرمال خارج می‌شود. نتیجه‌ی که از این قانون می‌توان گرفت این است که اشتغال جوامع با میزان بیماری‌هایی که برای افراد اتفاق خواهد افتاد ارتباط مستقیم دارد و دولت‌ها با تمرکز بیشتر بر اشتغال و حل این موضوع می‌توانند نقش مهم و تأثیرگذاری در بهبود سلامت جامعه داشته باشند. در قانون شماره ۱۰ به این نکته اشاره می‌شود که آن دسته از مناطقی که از نظر شاخص محرومیت و کمبود درآمد (IDACI)^۹ — که تأثیر مستقیمی بر تربیت، آموزش و... کودکان در وضعیت بحرانی دارد — و نیز از نظر اشتغال با مشکلات فراوان مواجه‌اند و از نظر محیط و شرایط اقلیمی نیز از وضع خوبی برخوردار نیستند، با اطمینان ۴۵٫۴۳ درصد اولین تأثیرشان بر ترک تحصیل افراد (از دانشگاه یا مدرسه) است.

قانون شماره ۱۱ تأکید دارد که آن دسته از مناطقی که از نظر شاخص عدم ورود به آموزش عالی (دانشگاه) وضع مناسبی ندارند و در رتبه‌های پایینی قرار دارند و از نظر درآمدی نیز جایگاه مناسبی ندارند و همچنین جزء مناطق متوسط و کم‌جمعیت هستند، با اطمینان ۶۴٫۷۰٪ بخش عمده‌ی آن‌ها از سال‌های بالقوه‌ی زندگی (جوانی) را در کسب علم، ثروت و... از دست داده‌اند. نکته‌ی قابل برداشت از این قانون این است که برای بهبود وضعیت کلی جامعه و فراهم کردن بسترهایی برای ایجاد شرایط مناسب برای افراد جامعه که بتوانند از سال‌های توانمندی زندگی خود نهایت استفاده را ببرند، تمرکز بر سطح درآمد جامعه، نرخ رشد جمعیت و در نهایت فراهم کردن شرایطی برای ورود آسان و بهتر افراد در آموزش عالی می‌تواند کمک بزرگی باشد. در قانون شماره ۱۲ نیز به این نکته اشاره می‌شود که آن دسته از مناطقی که از لحاظ شاخص اشتغال دچار ضعف‌اند و از نظر شاخص عدم ورود به آموزش عالی نیز وضعیت مناسبی ندارند و در رتبه‌های بالای این شاخص قرار می‌گیرند، و نیز از لحاظ جمعیتی جامعه جوان باشد، آنگاه با اطمینان ۶۳٫۱۱٪ آن مناطق از نظر شاخص بیماری و معلولیت وضعیت مطلوبی ندارند. این قانون بیان می‌دارد که جامعه‌ی جوانی که برای آن‌ها اشتغال و فرصت ورود به دانشگاه و ادامه تحصیل

پس از تشکیل شبکه براساس الگوریتم پیشنهادی، نوبت به کشف گره‌های مرکزی می‌رسد که بیشترین تأثیرگذاری را در شبکه براساس وزن کسب شده دارند. برای این کار باید تمام ۸۴۴ گره به‌دقت و براساس وزن هر پیوند بررسی و معین شود که کدام گره‌ها بیشترین خروجی وزن دار، و کدام گره‌ها بیشترین ورودی وزن دار را دارند. نتایج این بررسی در نرم‌افزار گفی عبارت است از:

۱. شاخص‌ها و رتبه‌های تأثیرگذار مرکزی‌ترین گره‌های شبکه. در جدول ۵ تأثیرگذارترین گره‌های شبکه مشخص شده‌اند. به‌عنوان مثال برای اولین گره مرکزی می‌توان چنین تفسیر کرد که شاخص درآمد جزء تأثیرگذارترین شاخص‌هاست و ۲۶۸ لینک خروجی به دیگر گره‌ها دارد. بنابراین اگر در جامعه‌ی بخواهیم شاخص محرومیت آن مناطق را بهبود ببخشیم ابتدا باید از شاخص درآمد شروع کنیم. همچنین رتبه‌ی ۶۱ بیان می‌دارد که باید در پایگاه داده خود جست‌وجو کنیم و ببینیم کدام مناطق و محدوده‌ها در اکثریت موارد از نظر شاخص درآمد رتبه‌ی ۶۱ را کسب کرده‌اند و اگر ما درآمد را در این نوع مناطق بهبود ببخشیم بیشترین بازدهی را بر بهبود دیگر شاخص‌های محرومیت خواهیم داشت.

۲. شاخص‌ها و رتبه‌های تأثیرپذیر مرکزی‌ترین گره‌های شبکه. در جدول ۶ نیز تأثیرپذیرترین گره‌های شبکه مشخص شده‌اند. به‌عنوان مثال شاخص بیماری و معلولیت در این شبکه ۲۳۸ لینک ورودی دارد و جزء شاخص‌های تأثیرپذیر به شمار می‌رود. این بدان معناست که این شاخص و مناطقی که در اکثریت موارد رتبه‌ی ۴۴ را کسب کرده‌اند یک وضعیت لب مرزی دارند. بنابراین اگر دیگر

جدول ۵. گره‌های تأثیرگذار در شبکه.

گره	تعداد پیوندهای ورودی
درآمد = ۶۱	۲۶۸
سال‌های بالقوه زندگی = ۳۴	۲۶۵
اشتغال = ۴۴	۲۶۱
اشتغال = ۴۷	۲۶۱
بیماری و معلولیت = ۲۷	۲۶۰
درآمد = ۶۳	۲۵۹
درآمد = ۶۲	۲۵۹
سال‌های بالقوه زندگی = ۲۴	۲۴۹
فاصله تا مدرسه ابتدایی = ۱۱	۲۳۴

جدول ۶. گره‌های تأثیرپذیر در شبکه.

گره	تعداد پیوندهای ورودی
درآمد = ۶۶	۲۶۸
مرگ و میر = ۴۸	۲۴۵
بیماری و معلولیت = ۴۴	۲۳۸
اشتغال = ۴۷	۲۳۸
ترک تحصیل = ۴۵	۲۳۷
درآمد = ۶۵	۲۳۵
سال‌های بالقوه زندگی = ۴۹	۲۳۰
جرم و جنایت = ۲۹	۲۲۴
درآمد = ۶۳	۲۲۴

شاخص‌هایی را که در قسمت قوانین انجمنی به دست آوردیم و تأثیر مستقیمی بر این شاخص دارند ارتقا بدهیم، این شاخص به سادگی از این حالت بد خارج شده و مقاومتی نشان نخواهد داد و به وضع بهتری دست خواهد یافت.

۶. نتیجه‌گیری

کشف روابط پنهان در میان شاخص‌های محرومیت و کشف روابط تأثیرگذار در نیازها و کمبودهای اقشار مناطق مختلف می‌تواند نقش بسیار تأثیرگذاری در محرومیت‌زدایی هدف‌مند از مناطق محروم و تخصیص بودجه‌ی مناسب برای این کار داشته باشد. در این مقاله سعی شده تا با استفاده از فنون داده‌کاوی به تجزیه و تحلیل شاخص‌های محرومیت پرداخته و با استخراج دانش‌های پنهان در میان آن‌ها به نتایجی در راستای بهبود هرچه بهتر این شاخص‌ها و محرومیت‌زدایی از مناطق کم‌تر توسعه‌یافته، دست یابیم. همچنین با کشف تأثیرگذارترین و تأثیرپذیرترین شاخص‌ها توسط شبکه‌ی پیچیده‌ی طراحی شده، اطلاعات این شبکه تجزیه و تحلیل و از نتایج آن در تصمیم‌گیری‌های محرومیت‌زدایی استفاده خواهد شد.

نتایج به دست آمده از این تحقیق ما را در کشف روابط پنهان و تأثیرگذاری‌های پنهان در میان کمبودها و نیازهای اقشار مختلف جامعه، تصمیم‌گیری‌های مناسب‌تر برای رفع این کمبودها، اختصاص بودجه و تمرکزگرایی در محرومیت‌زدایی یاری می‌دهد. این تصمیم‌گیری‌ها در گذشته به‌صورت تجربی و براساس گزارشات آماری استخراج شده از سوی شاخص‌های محرومیت انجام می‌گرفت و به تأثیرگذاری‌ها و روابط پنهان در میان این شاخص‌ها توجهی نمی‌شد. به‌عنوان نمونه‌ی این نتایج این تحقیق می‌توان به این موضوع اشاره کرد که تمرکز بر سطح درآمدی مناطق، جمعیت آن‌ها، مهارت‌های آموزشی و ایجاد شرایط برای استفاده بهتر از سال‌های بالقوه زندگی، تأثیر بسیار زیادی بر کاهش احتمال وقوع جرم و جنایت در مناطق جرم‌خیز دارد. تصمیم‌گیران با تمرکز بر این متغیرها می‌توانند با صرف هزینه و انرژی کم‌تری از وقوع جرم و جنایت در این‌گونه مناطق بکاهند. دیگر قوانین استخراج شده نیز به‌همین ترتیب به این نوع تصمیم‌گیری‌ها کمک می‌کنند.

همچنین نتایج استخراج شده از بخش شبکه‌ی پیچیده‌ی قوانین و کشف تأثیرگذارترین و تأثیرپذیرترین شاخص‌ها و متغیرها در جوامع، می‌تواند در برنامه‌های کوتاه‌مدت با بازدهی بالا در محرومیت‌زدایی و همچنین در برنامه‌های بلندمدت برای محرومیت‌زدایی از مناطق محروم در سطوح بزرگ کشوری مورد استفاده قرار گیرد.

نتایج به دست آمده از این تحقیق مربوط به دانش‌ها و الگوهای پنهان موجود در محرومیت کشور انگلستان است که با توجه به پیشرفته بودن این کشور، استفاده از این الگوهای کشف شده برای کشورهای درحال توسعه همچون ایران می‌تواند مفید واقع شود. همچنین در تحقیقات آتی پژوهش‌گران می‌توانند با استفاده از روش‌های پیش‌بینانه‌ی داده‌کاوی براساس خواسته‌ها و نیازهای شهروندان و با ارائه‌ی وزن‌هایی براساس فرهنگ، درآمد ملی، سطح پیشرفته بودن و... این نتایج را به دیگر کشورها تعمیم بدهند.

۷. محدودیت‌های تحقیق و تحقیقات آتی

نتایج به دست آمده از این تحقیق شامل الگوهای پنهان موجود در شاخص‌های محرومیت کشور انگلستان است و برای دیگر کشورها به‌طور مستقیم قابل تعمیم نیست. اما روش ارائه شده در این تحقیق یک روش نوین در این حوزه است که

در خصوص تعمیم نتایج می‌توان در تحقیقات آتی با محاسبه‌ی این شاخص‌ها یا مشخصه‌های مشابه، با توجه به خواسته‌ها و نیازهای شهروندان و با ارائه‌ی وزن‌هایی براساس فرهنگ کشور مورد نظر، درآمد ملی، سطح پیشرفته بودن و... قوانین به دست آمده را آزمود و تعمیم‌پذیری این نتایج را مورد بررسی قرار داد.

افقی جدید در زمینه‌ی استفاده از فنون داده‌کاری و شبکه‌های پیچیده در این حوزه را باز کرده و نشان می‌دهد که این فنون در این حوزه مؤثرند و ما را به سمت قوانین و دانش‌هایی رهنمون می‌سازد که بتواند در زمینه‌ی تمرکزگرایی در محرومیت‌زدایی مفید واقع شود.

پانویس‌ها

1. indices of multiple deprivation
2. support
3. confidence
4. human development index
5. indices of multiple deprivation
6. cross industry standard process for data mining (CRISP)
7. supervised
8. unsupervised
9. income deprivation affecting children people index

منابع (References)

1. Manzari, M. and Mohgher, A. "Applying a model of credit allocation of deprivation", *Journal of Planning and Budget*, **91**, pp.69-90 (2011).
2. Heydari, R. and Farajolahi, M. "Providing a new index to assess urban-rural deprivation", *Journal of Urban Economics*, **8**, pp.18-20 (2010).
3. Rezvani, M., Mansorian, A. and Satari, H. "Evaluation and development of indicators of urban life quality (Case Study: Nur Abad, Lorestan)", *Journal of Urban and Regional Studies and Research*, **2**, pp.87-110 (2009).
4. Omrani, M., Pajoyan, J. and Sharifzadeh, M. "Evaluation of poverty in Tehran with using the inverse of the median and average food expenditure Engel coefficient", Islamic Azad University, Science & Research (2011).
5. Chapman, J., Congdon, P. and Shaw, S.Y.H. "Carter the geographical distribution of specialists in public health in the United Kingdom: Is capacity related to need?", *Journal of The Royal Institute of Health, Public Health*, **119**, pp. 639-464 (2005).
6. Gartner, A., Farewell, D., Roach, P. and Dunstan, F. "Rural/urban mortality differences in England and Wales and the effect of deprivation adjustment", *Social Science & Medicine*, **72**, pp. 1685-1694 (2011).
7. Catherine, A., Harrison, D.A., Hutchings, A. and Kathryn, R. "The association between deprivation and hospital mortality for admissions to critical care units in England", *Journal of Critical Care*, **25**, pp. 382-390 (2010).
8. Giovanni, F., Yves, F. and Sylvain, W. "A cluster analysis of multidimensional poverty in Switzerland", Centre de Recherche Appliquee en Gestion (2006).
9. Kuznetsov, L., Maier, W., Hunger, M., Meyer, M. and Mielck, A. "Associations between regional socioeconomic deprivation and cancer risk: Analysis of population-based cancer registry data from Bavaria, Germany", *Preventive Medicine*, **53**, pp. 328-330 (2011).
10. Han, J. and Kamber, M., *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufman Publisher (2006).
11. Russel, P., Sing Koh, Y., Dobbie, G. and Yeap, W. "Weighted association rule mining via a graph based connectivity model", *Information Sciences*, **218**, pp. 61-84 (2013).
12. Minaei-Bidgoli, B. and Akhondzadeh, E. "A new approach of using association rule mining in customer complaint management", *IJCSI International Journal of Computer Science Issues*, **7**(5), ISSN (Online), pp. 1694-0814 (2010).
13. Newman, M.E.J. "The structure and function of complex networks", *Society for Industrial and Applied Mathematics, SIAM REVIEW*, **45**(2), pp. 167-256 (2003).
14. Xiaobing, L., Zhai, K. and Pedrycz, W. "An improved association rules mining method", *Expert Systems with Applications*, **39**, pp. 1362-1374 (2012).
15. Mir Mohamad Sadeghi, M., *Social Network Analysis with NodeXL*, Tehran, Kian Publications (1391).
16. Taghvai, M. and Abdolahi, A. "Cluster analysis of the status of development and deprivation of countries with using the human development index", *Journal of Geographical Research*, **93**, pp.129-148 (2009).
17. Duckworth, A.D., Clement, N.D., Jenkins, P.J., Will Charles, E.M., and McQueen, M.M. "Socioeconomic deprivation predicts outcome following radial head and neck fractures", *Injury, Int. J. Care Injured*, **43**, pp. 1102-1106 (2012).
18. Koenig, A., Samarasundera, E. and Cheng, T. "Interactive map communication: Pilot study of the visual perceptions and preferences of public health practitioners", *Public Health*, **125**, pp. 554-560 (2011).

19. Alafar, A. "Fighting poverty by creating employment", *Study of Poverty and Poverty Reduction, Plan and Budget Organization*, **2**, pp.79-100 (2010).
20. Chapman, J., Congdon, P., Shaw, S. and Carter, Y.H. "The geographical distribution of specialists in public health in the United Kingdom: Is capacity related to need?", *Journal of the royal institute of health, Public Health*, **119**, pp. 639-464 (2005).
21. Victoria, P.A. and Jacinto, M.V. "An evolutionary algorithm to discover quantitative association rules from huge databases without the need for an a priori discretization", *Expert Systems with Applications*, **39**, pp. 585-593 (2012).
22. Ferreira, J. and Figueiredo, A.T. "An unsupervised approach to feature discretization and selection", *Pattern Recognition*, **45**, pp. 3048-3060 (2012).