



## کشف کیفیت در ماهیت و آموزش جغرافیا (QDGE) با استفاده از قواعد تلازم

دکتر ابراهیم مقیمی

استاد دانشگاه تهران

Moghimi\_ir@yahoo.com

### چکیده

کشف کیفیت همچون دانش دیگری با رویکرد کاوشی می‌کاود تا کیفیت را بدست آورد. کشف کیفیت ماهیتی جدیدی از دانش آوری را در بردارد. روشها و ابزارهای آن ماهیت اکتشافی دارد و به فراخور نیاز رشته‌ها و تخصصها فراهم می‌شود. دانش جغرافیا به مانند سایر دانشها در ماهیت و آموزش نیازمند رویکرد اکتشافی به کیفیت است تا بتواند بالندگی خود، محیط و جامعه را استمرار دهد. کشف کیفیت در ماهیت و آموزش رشته جغرافیا (QDGE) در واقع سراجی خواهد بود که نقشه راه را در آینده نشان خواهد داد و متضمن حقایق و روابطی است که متقاعد ساز ذهن است و فراگیری، آموزش، و استفاده را ضروری می‌سازد، و براین اساس نیازمند جستار شناسی از ذهن پذیری فردی، گروههای علمی و انجمنها از یک طرف و از طرف دیگر ماهیت، رشته‌های وابسته، دانش آموختگان و حرفه پذیری در آموزش جغرافیا است که در این مقاله با توجه به مستندات موجود در ایران از جمله چشم‌انداز جامعه کوچک و جدید رشته‌های جغرافیا، با استفاده از قواعد تلازمی آن را کیفیت شناسی خواهیم کرد.

کلید واژه ها: کشف، کیفیت، رشته جغرافیا، قواعد تلازمی

## مقدمه

ایده کشف کیفیت (Quality Discovery) به عنوان یک دانش جدید میان رشته‌ای، در حال رشد است و حوزه‌های مختلفی دانشی از جمله جغرافیا را شامل می‌شود. پایگاه داده، آمار، یادگیری، تحلیل و سایر زمینه‌های مرتبط را با هم تلفیق و کیفیت ارزشمند نهفته در حجم بزرگی از داده‌ها را استخراج می‌نماید. همانطور که فرکانس الکترونها و امواج موضوع اصلی کیفیت مهندسی صدا هستند نتایج حاصل از تحلیل داده‌ها، رشته‌ها، گرایشها، مدرسین، اطلاعات دانشجو، آموزش و توسعه نیز موضوع اساسی کیفیت در ماهیت آموزش از جمله آموزش جغرافیا (Quality Discovery Geography Education) می‌باشد که به اختصار QDGE می‌نامیم.

به طور کلی کیفیت را رشته‌ای از بیتهای (به صورت صفر تا یک) یا اعداد و نشانه‌ها و یا اشیا هستند که وقتی در فرمتی مشخص به یک برنامه ارسال می‌شوند، معنا می‌یابند (غضنفری و همکاران، ۱۳۸۷). کیفیت در واقع به عنوان تصاویر ذهنی و عینی، درک، کشف یا فراگیری می‌شود. محققانی که از حوزه‌های مختلف به کیفیت می‌پردازند ممکن است فهم متفاوتی از عبارات «کشف کیفیت» و کیفیت داشته باشند. کشف کیفیت از پایگاه داده‌ها در واقع فرایند تشخیص الگوها (patterns) و مدل‌های موجود در داده‌هاست (Fayyad and at all, 1996). نتایجی که معتبر، بدیع (Novel) بالقوه مفید و کاملاً قابل فهم هستند کیفیت برترند. هدف کشف کیفیت یافتن رضایتمندی و نتایج مطلوب موجود در داده‌ها است که در میان حجم عظیمی از داده‌ها و اقدامات تصدیقی پنهان است. تصدیق یا گواهی (Evidence) نشانگر معنی دار بودن یک «یافته» بر حسب یک معیار آماری یا ارزشی است (مقیم و محمودی، ۱۳۸۳). مطلوبیت و برتری، مقدار شباهت یک الگوی کشف شده نسبت به یافته‌های دیگر است و درجه تبعیت آنرا از دیگری اندازه می‌گیرد. رضایتمندی، ارزش ارتباط یافته را با اهداف کاربران بیان می‌کند. بدیع بودن (Novelty) بیانگر میزان تازگی نسبت به دانش قبلی کاربر یا کاشف سیستم است. سادگی (simplicity) به پیچیدگی نحوی (Syntactical) و نمایش یک الگوی کشف شده و نحوه تعمیم آن اشاره دارد. از این رو با کشف دانش متفاوت است.

## ۲- بحث و تحلیل

### ۱-۲) کشف دانش در پایگاه داده‌ها (KDD) و داده کاوی (DD)

از دیدگاه منطقی، «دانش» حقیقت صریحاً اظهار شده و موجه در یک زمینه است که با زبانهای تخصصی بیان می‌شود. کشف دانش، گزاره‌هایی را تولید می‌کند که حقیقت اشیاء جهان، مفاهیم و نظمها را توصیف و شفاف می‌کنند. پایگاه‌های داده، مخازنی از داده‌ها درباره زمینه‌های مختلف دنیای واقعی می‌باشند. KDD بیش از تحلیل داده‌ها و فراتر از کشف الگو است. بسیاری از الگوهای موجود در داده‌ها، ماهیت دانشی ندارند (Klasgen and at all, 2002).

شاپیرو که در سال ۱۹۸۹ واژه KDD را ابداع کرده است می‌گوید: «واژه KDD ابتدا در جامعه دارای هوش مصنوعی و یا ابزاری (صنعتی) متداول شد. ولیکن محققان پایگاه داده‌ها که در ارتباط بیشتری برای گفتمان با اهل فن و رسانه‌ها بودند و داده‌هایی را در اختیار داشتند، بیشتر استفاده کردند. داده کاوی (DD) واژه‌ای قدیمی تر از KDD است و در جامعه تحلیل داده‌های آمار محور، ابداع شده است (Shapiro, 2000). فیباید و همکاران استفاده کننده از داده کاوی را واژه‌ای همراه کننده می‌دانند. آنها معتقدند که در واقع آهن و طلا کاوش می‌شوند، کاوش از خود طلا یا آهن است نه غبار یا سنگ و خاک. لذا دانش کاوی و داده کاوی دو مقوله متفاوت است. دانش کاوی مانند آهن و طلا، خروجی مورد نظر، دانش است، نه طلا. برخی، داده کاوی را یک گام مرکزی در فرایند کشف دانش می‌دانند که الگوریتمهای استخراج و اثبات فرضیه را اعمال می‌کند. بسیاری داده کاوی را معادل با واژه متداول کشف دانش در پایگاه داده‌ها می‌دانند (Fayyad and at all, 1996). عناوین دیگری که در گذشته به جای داده کاوی استفاده می‌شدند عبارتند از: باستان شناسی داده (Data Archaeology)، لایروبی داده (Data Dredging)، تحلیل وابستگی تابعی و درو کردن داده (Data Harvesting). همه اینها ماهیت بررسی دارند ولی ماهیت اکتشافی دانشی ندارند (Pregibon, 1999).

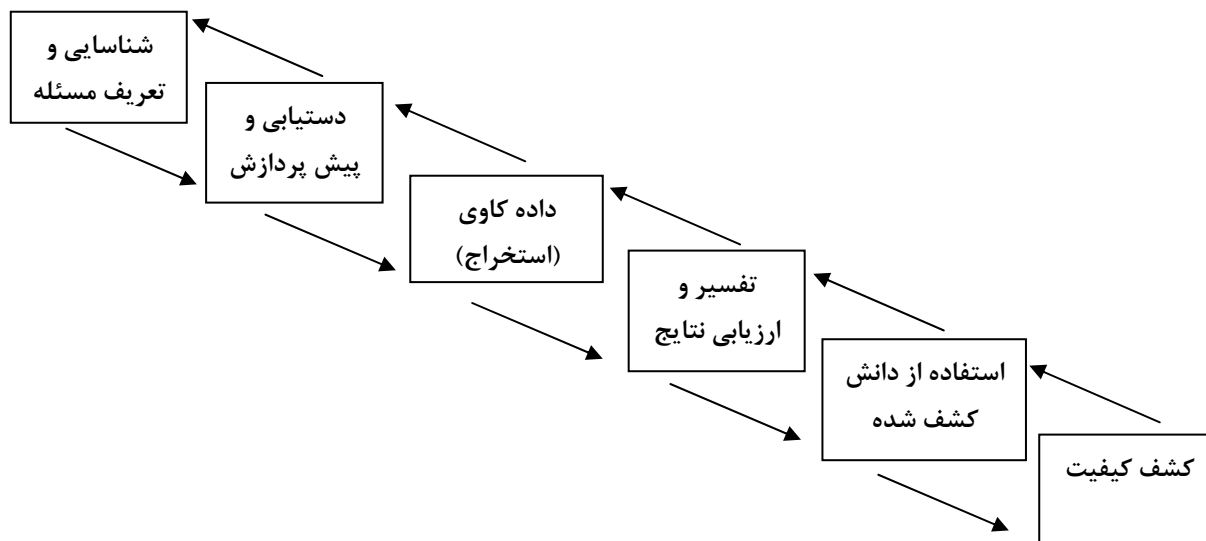
باید توجه داشت که در زبان فارسی فعل «کاویدن» هم برای داده‌ها و هم برای دانشی که از داده‌ها استخراج می‌شود، قابل استفاده است. یعنی اصطلاح کاویدن بر روی داده‌ها و کاویدن بر روی دانش، هر دو درست است. ولی ما در این مقاله؛ با تفکیک مستقل استفاده می‌کنیم.

یعنی واژه «کاویدن» را برای داده‌ها و برای دانش از واژه «کشف» استفاده کرده و واژگان داده کاوی و کشف دانش را به کار می‌بریم. و هر دو را زمینه و اساس کشف کیفیت می‌دانیم.

### ۳-۱) فرایند کشف کیفیت

بر اساس دیدگاه و مطالب فوق که کشف دانش و داده کاوی را بخشی از فرایند کشف کیفیت می‌دانند، چرخه کشف کیفیت شامل مراحل متعددی مطابق با شکل یک است و متدولوژی آن در شکل دو آمده است. طبق شکل شماره یک اولین مرحله: شناسایی و تعریف مسئله است.

مرحله دوم: جمع آوری و پیش پردازش داده‌ها (Preprocess) شامل انتخاب داده‌ها، حذف نقاط پرت (Quilters) یا مغشوش (Noise)، طرز برخورد با داده‌های مفقوده (Missing Data) و تبدیل (Transformation) و یا گسسته سازی (Discrimination) و کاهش (Reduction) داده‌ها است. این مرحله معمولاً در کل فرایند بیشترین زمان را می‌برد.



شکل ۱) سیکل یا چرخه کشف کیفیت (اقتباس از برری و لینف، ۲۰۰۴، و غضنفری و همکاران، ۱۳۸۷، با تغییرات).

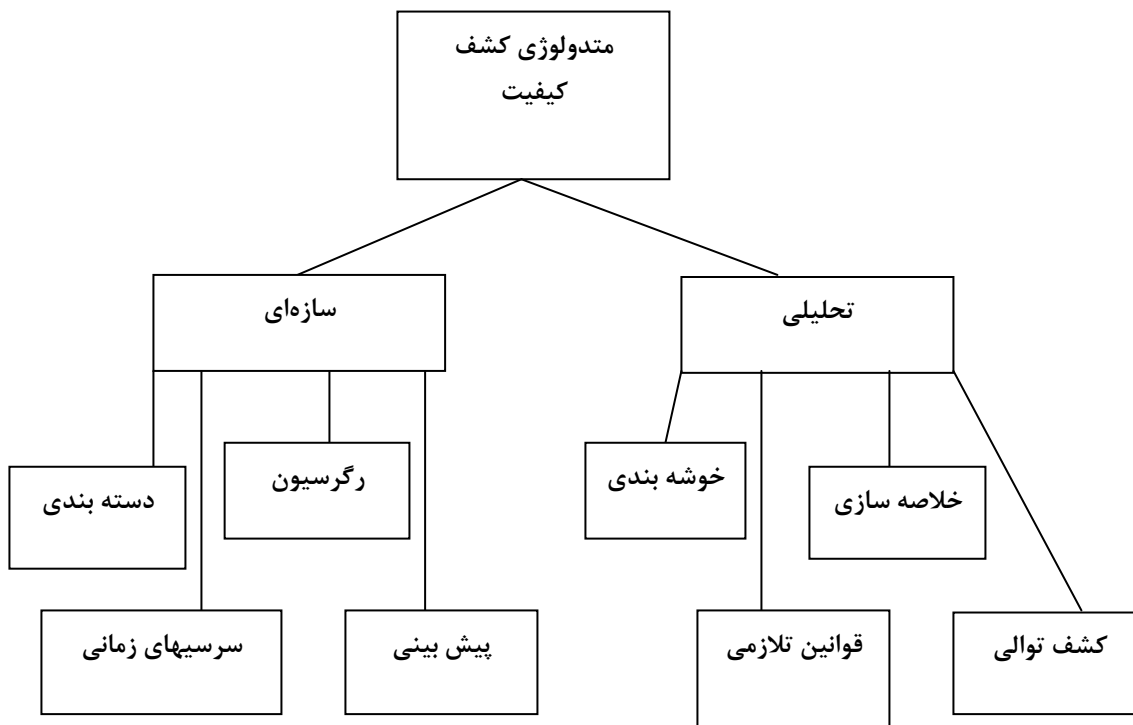
مرحله سوم: داده کاوی است که هدف آن استخراج الگوها و یا مدل‌های مخفی یا پنهان کرانه دار (حد اکثری یا حداقلی) در داده‌ها است. روش‌های اصلی داده کاوی به دو دسته توصیفی (Descriptive) و پیش‌بینانه تقسیم می‌شوند. نمونه‌هایی از این روش‌ها عبارتند از: مدل‌سازی برای پیش‌بینی (Predictive) (مثل دسته‌بندی (Classification) و رگرسیون (Regression))، طبقه‌بندی یا تقطیع (Segmentation) (خوشه‌بندی (Clustering))، مدل‌سازی وابستگی (Dependency) (مانند مدل‌های تصویری یا تخمین چگالی)، تلخیص (خلاصه‌سازی) (Summarization) پیدا کردن رابطه بین فیلدها، تلازم یا انجمنی، مصور سازی (Visualization) و مدل‌سازی یافتن تغییر و انحراف (Deviation) در داده و دانش. مرحله چهارم: تفسیر (یا فرا پردازش (post-process)) داده کشف شده است. این

تفسیر، عملاً توصیفی یا پیشبینانه است که دو هدف اصلی سیستم‌های اکتشافی می‌باشند. تجربه نشان داده است که همیشه الگوها یا مدل‌های کشف شده از داده‌ها، مفید و جالب نیستند.

مرحله پنجم: کشف دانش است، گزاره‌ای که حقیقت، مفاهیم و نظم وجودی داده را به طور انحصاری و برای اولین بار برای محقق ایجاد و فراهم می‌کند. مرحله ششم: استفاده عملی از دانش کشف شده است. برخی اوقات می‌توان از دانش کشف شده بدون آنالیز کردن آن استفاده کرد. در مواقع دیگر کاربر انتظار دارد دانش کشف شده از طریق یک برنامه آنالیز شده به کار گرفته شود (غضنفری و همکاران، ۱۳۸۷).

مرحله هفتم: به کارگیری قطعی نتایج اثر بخش داده کشف شده، متناسب با هدف‌های ارزشی، بینشی و دانشی، و سرانجام تایید یا رد آن هدف است و این فرایند نهایی است.

مجموعه این فرایندها در عین حال که هر کدام خود نیازمند کیفیت هستند، به کشف کیفیت منتهی می‌شود. توجه کنید که کشف کیفیت شامل ارزیابی و احتمالاً تفسیر الگوها برای تفکیک خوب از غیر خوب است و متدولوژی آن در شکل دو آمده است.



شکل ۲) متدولوژی کشف کیفیت (اقتباس از برری و لینف، ۲۰۰۴، با تغییرات)



#### ۱-۴) چرا QDGE لازم است؟

جغرافیا دانش میان رشته ای است (هاگت، ۱۳۷۵) و نیاز مند کیفیت شناسی در سطح ملی و جهانی است. برخی از دلایل نیاز به QDGE به شرح زیر می باشد پاسخ به سوالات زیر یعنی ضرورت توجه به QDGE :

۱. بسیاری از سازمانها و دانشگاه هایی که کارهای جغرافیایی انجام می دهند، داده‌های زیادی جمع کرده اند، با آن چه می کنند؟
۲. محققین دانشجویان و کاربران جغرافیا ممکن است همه داده ها را ذخیره کنند زیرا فکر می کنند آنها بطور ضمنی حاوی دارایی با ارزشی هستند.
۳. آیا داده های در اختیار، عملکرد بهتری را برای بهبود فرایند و رفع مشکلات فراهم می آورند. چگونه با آن داده‌ها تصمیم می گیرند، تصمیم صحیح کدام است؟
۴. آیا همه‌ی داده های جمع آوری شده تحلیل پذیرند؟ با توجه به نرخ رشد داده ها، به کارگیری روشهای سنتی ( که دستی و زمان بر هستند) برای تحلیل آنها کار ساز و اشتباه آمیز خواهد بود؟
۵. آیا داده هایی که ممکن است هرگز تحلیل نشوند و با هزینه زیاد و بطور پیوسته جمع آوری شده قابل تفکیک از دیگر داده ها هستند. چگونه می توان میلیونها رکورد و دهها یا صدها فیلد را کاوش کرد و کیفیت را یافت؟
۶. آیا شبکه سازی، فرصتی مناسب و رو به رو رشد برای دسترسی بیشتر به QDGE را فراهم می کند؟ چه شبکه جغرافیایی برای کشور و جهان متصور است؟
۷. آیا QDGE می تواند نیاز به تشخیص و پاسخ سریع به فرصتهای در حال ظهور را فراهم کند. چه مکانیسمی برای این کار لازم است؟
۸. آیا QDGE اطلاعات کسب و کار و مشاغل جغرافیایی مطلوب را به ما می دهد؟ چه مصداقی برای این منظور موجود است؟
۹. آیا QDGE مهارت شغلی برای دانش آموختگان خود فراهم می کند؟ چه نسبتی از دانش آموختگان فاقد مهارت شغلی می باشند؟
۱۰. آیا QDGE امکان تحلیل داده های آموزشی مورد نیاز را برای ما فراهم می کند؟ چه نوع داده های مستندی در کتابها و منابع جغرافیایی منتخب برای آموزش وجود دارد؟
۱۱. آیا QDGE راه توسعه رشته ها و آموزش دانشجویان را برای ما هموار می کند؟ چه نوع رشته هایی پیشگام توسعه آموزش جغرافیا هستند؟ مثلا در ژنو مورفولوژی چه نوع گرایشهایی پیشگام است (مقیمی، ۱۳۸۹، ب)
۱۲. آیا QDGE امکان سرمایه گذاری برای توسعه دانش جغرافیا را به ما می دهد؟ چه نوع سرمایه گذاری های فیزیکی و انسانی برای آموزش جغرافیا در آینده لازم است؟
۱۳. آیا QDGE امکان تشخیص مسیر و سبیل نادرست را برای ما روشن می کند؟ نقشه راه برای آموزش جغرافیا کدام است؟



۱۴. داده های جغرافیایی چگونه تولید می شوند و یا از کجا می آیند؟ چه کسی یا کسانی آن‌ها را تولید می کنند؟ چه کسانی باید آنها را تولید کنند (هاگت، ۱۳۷۵)؟

۱۵. زمان تولید داده ها چه وقت است؟ در آینده چه نوع داده‌هایی تولید خواهند شد؟ انتظار ما از آینده چیست؟ مثلاً چه نوع انتظاری از تو سعه شهر ها در آینده متصور است (مقیمی، ۱۳۸۹، الف)؟

موارد فوق از جمله مواردی است که پرداختن به QDGE و تحلیل آن را ضروری می سازد.

به نظر این‌جانب، یکی از کاربردی ترین حالت های استفاده از تحلیل قواعد تلازمی، تجزیه و تحلیل کیفیت یک رشته علمی است. از جمله با استفاده از قواعد تلازمی می‌توان رشته های جغرافیا را تحلیل کرد. تحلیل تلازمی، مطالعه ویژگیها یا داده هایی می باشد که با یکدیگر همراه بوده. این روش به دنبال استخراج قواعد به منظور کمی کردن ارتباط میان دو یا چند داده است. قواعد تلازمی به شکل اگر و آنگاه به همراه دو معیار پشتیبان و اطمینان (Confidence) تعریف می شوند. پیشرفت فناوری دانشگاه ها را قادر می سازد تا حجم زیادی از داده های دانشجویان، اعضای هیات علمی و دانش آموختگان (که در این مقاله از آن به عنوان سبد دارایی یک رشته علمی یاد می شود) را جمع آوری و ذخیره نمایند. هر رشته رشد و توسعه متفاوتی را در مقادیر مختلف و زمانهای متفاوت انجام می دهد و داده های موجود در سبد نشان دهنده رشد و توسعه در یک زمان خاص است. هیات علمی، کتابها و مقالات، تنوع رشته‌ها، مشاغل، درجه حساسیت، ساختار، نهاد و... از جمله داده-هایی است که می توان آنها را در سبد جغرافیا قرار داد. با تجزیه و تحلیل سبد می توانند رفتار صنفی علمی رشته جغرافیا را پیش بینی کرد. این کار به جغرافیدانان کمک می کند تا بتوانند رشته خود را بهتر ساماندهی و چیدمان بهتری از رشته خود داشته باشند و از این طریق ماندگاری خود را افزایش دهند. تحلیل تلازمی تنوع گرایش‌های رشته جغرافیا به شرح زیر است:

- بررسی این که چه گرایشها و رشته‌هایی از جغرافیا در یک دانشگاه با یکدیگر ارائه می شوند و اینکه چه گرایشها و رشته هایی هیچ گاه ارائه نمی شوند.

- تعیین سهم گرایشها و رشته های فعال و بررسی تاثیرات آن‌ها در ایجاد یک رشته جدید از جغرافیا.

قواعد تلازمی ماهیتاً قواعد احتمالی هستند. به عبارت دیگر قاعده  $X \Rightarrow A$  برای مجموعه علوم جغرافیایی (I) لزوماً قاعده  $X + Y \Rightarrow A$  را نتیجه نمی دهد، زیرا این قاعده ممکن است از شرط حداقلی پشتیبان برخوردار نباشد. به طور مشابه قواعد  $X \Rightarrow Y, X \Rightarrow Z$  لزوماً قاعده  $X \Rightarrow Z$  را نتیجه نمی دهند زیرا قاعده اخیر ممکن است از شرط حداقلی اطمینان برخوردار نباشد.

نمای کلی یک قاعده تلازمی به شکل زیر می باشد:

اطمینان، پشتیبان  $X \Rightarrow Y$

به طور کلی داریم:

$X \subset I, Y \subset I, X \cap Y = \Phi$

- پشتیبان (Support)  $(X, Y)$ : نشان دهنده درصد یا تعداد مجموعه رشته ها است که شامل هر دوی X و Y باشند.

- اطمینان (Confidence): میزان وابستگی یک رشته خاص را به دیگری بیان می کند و مطابق فرمول زیر محاسبه می شود:



$$(X) \text{ پشنتیبان } / (X \cup Y) \text{ پشنتیبان } = (X, Y) \text{ اطمینان}$$

این شاخص درجه وابستگی بین دو مجموعه  $X$  و  $Y$  را محاسبه میکند و به عنوان شاخصی برای اندازه گیری توان یک قاعده در نظر گرفته می شود. غالباً قاعده ای انتخاب می شوند که عدد اطمینان بزرگی داشته باشند.

فرض کنید اطلاعات دانشگاه‌هایی که رشته  $X$  را دایر کرده اند همچنین علاقه دارند در همان زمان رشته  $Y$  نیز دایر کنند در قاعده تلازمی زیر نشان داده شده است.

$$X \Rightarrow Y \quad (\text{پشنتیبان} = 20\% \text{ و اطمینان} = 60\%)$$

شاخصهای اطمینان و پشنتیبان قواعد بیانگر جذابیت آنها هستند این دو شاخص به ترتیب مفید بودن و اطمینان از قواعد مکشوفه را نشان می دهند. پشنتیبان 20٪ برای قاعده تلازمی فوق به این معنی است که 20٪ همه داده های موجود نشان می دهند که رتبه  $X$  و  $Y$  با هم مورد استقبال قرار گرفته اند. اطمینان 60٪ به این معنی است که 60٪ دانشگاه‌هایی که رشته  $X$  را دایر کرده اند رشته  $Y$  را نیز دایر کرده اند.

لذا با استفاده از قواعد تلازمی فوق و سبد مجموعه علوم جغرافیایی، به کشف می پردازیم مجموعه، علوم جغرافیایی را به صورت زیر فرض می کنیم:

$$I = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

$I = \{1 - \text{اقلیم شناسی، 2 - ژئومورفولوژی 3 - جغرافیای شهری، 4 - جغرافیای روستایی، 5 - جغرافیای سیاسی، 6 - گردشگری، 7 - مخاطرات محیطی، 8 - سنجش از راه دور و GIS}\}$

مجموعه دانشگاه ها ( $D$ ) و ( $U$ ) دانشگاه خاص است.

$$D = \{U_1, U_2, U_3, U_4, U_5, U_6, U_7, U_8\}$$

$$U_1 = \{\text{سنجش از راه دور، مخاطرات، جغرافیای روستایی، جغرافیای شهری، ژئومورفولوژی}\}$$

$$U_2 = \{\text{مخاطرات، جغرافیای شهری، ژئومورفولوژی}\}$$

$$U_3 = \{\text{جغرافیای سیاسی، گردشگری، مخاطرات، جغرافیای شهری، ژئومورفولوژی}\}$$

$$U_4 = \{\text{جغرافیای سیاسی، گردشگری، مخاطرات، ژئومورفولوژی}\}$$

$$U_5 = \{\text{جغرافیای سیاسی، گردشگری، اقلیم شناسی}\}$$

$$U_6 = \{\text{اقلیم شناسی، جغرافیای سیاسی}\}$$

$$U_7 = \{\text{ژئومورفولوژی، اقلیم شناسی، سنجش از راه دور}\}$$



$$U_8 = \{\text{سنجش از راه دور، جغرافیای سیاسی}\}$$

فرض کنیم یک قاعده تلازمی به شکل زیر داریم:

$$X \Rightarrow Y = [\text{اطمینان، پشتیبان}]$$

{ژئومورفولوژی، سنجش از دور، جغرافیای شهری، جغرافیای روستایی}

$$X = \{\text{جغرافیای شهری، جغرافیای روستایی}\}$$

$$Y = \{\text{ژئومورفولوژی، سنجش از دور}\}$$

$$X \cup Y = \{\text{جغرافیای شهری، جغرافیای روستایی، ژئومورفولوژی، سنجش از دور}\}$$

$$\text{پشتیبان} \quad (X \cup Y) = \frac{2}{8} = 0.25$$

از آنجایی که مجموعه  $X \cup Y$ ، ۲ عضو و مجموعه  $D$ ، ۸ عضو دارد، بنابراین الگوی «جغرافیای شهری، جغرافیای روستایی، ژئومورفولوژی، سنجش از دور» در ۲۵٪ سبد ما رخ می دهد.

یعنی جغرافیای شهری، جغرافیای روستایی، در  $U_1, U_2, U_3, U_4, U_5, U_6, U_7$  استقبال شده اند.

(۳) نتیجه

بنابراین داریم:

$$\text{پشتیبان} \quad (x) \frac{5}{8} = 62\%$$

$$\text{اطمینان} = \text{پشتیبان} \quad (X \cup Y) / (X) = \left(\frac{2}{8}\right) / \left(\frac{5}{8}\right) = 2/5 = 40\%$$

یعنی دانشگاه های ایران یافته اند که به جغرافیای شهری و جغرافیای روستایی نسبت به دیگر رشته ها اولویت بیشتری بدهند. از طرفی هنگامی که دانشگاه های ایران جغرافیای شهری، جغرافیای روستایی، دایر می کنند در ۴۰٪ اوقات، ژئومورفولوژی، سنجش از دور را نیز دایر میکنند. پس پیدا کردن تلازم و یافتن چنین مفاهیم محکم و قابل توجهی هدف اصلی کشف کیفیت برای رشته های جغرافیا ست.

(۴) منابع

1. Berry M.J.A. Linoff G.S.(2004) Data Mining techniques for Marketing, Sales, and Customer Relationship Management (2<sup>nd</sup> end), Wiley.





پنجمین همایش "ارزیابی کیفیت در نظام داشگاهی" دانشگاه تهران- پردیس دانشکده‌های فنی- اردیبهشت ماه ۱۳۹۰

2. Fayyad U., Piatetsky – Shapiro G., Smyth P., Uthurusamy R(1996) Advances in Knowledge Discovery and Data Mining MIT Press.
3. Klogsen W. and Zytkow J.M. (2002) Handbook of Data Mining and Knowledge discovery, Oxford university press.
4. Pregibon D (1999) "2001: a Statistics odyssey" Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and Data Mining.
5. Shapiro G.P. (2000) Knowledge discovery in Databases: 10 years after ACM SIGKDD Explorations, Feb 2000, Vol1, No 2.
۶. هاگت، پیتر، جغرافیا ترکیبی نو، ترجمه مرحوم شاپور گودرزی نژاد، دو جلد، انتشارات سمت، ۱۳۷۵.
۷. غضنفری، مهدی، سمیه علیزاده، بابک تیمور پور، داده کاوی و کشف دانش، دانشگاه علم و صنعت، ۱۳۸۷.
۸. مقیمی، ابراهیم و فرج اله محمودی، روش تحقیق در جغرافیای طبیعی، انتشارات قومس، ۱۳۸۳.
۹. مقیمی ابراهیم، ژئومورفولوژی شهری، انتشارات دانشگاه تهران، (الف) ۱۳۸۹.
۱۰. مقیمی، ابراهیم، ژئومورفولوژی ایران، انتشارات دانشگاه تهران (ب) ۱۳۸۹.
۱۱. یکی از مراجع اصلی اطلاعات درباره کشف دانش و داده کاوی سایت <http://www.kdnuggets.com> است که دارای خبرنامه ماهانه نیز می باشد.