تکنیک داده کاوی و کاربرد آن در مطالعات اجتماعی

گلمراد مرادی (عضو هیأت علمی دانشگاه آزاد واحد اسلام آبادغرب، نویسنده مسئول)
moradi.pop@gmail.com
وحید قاسمی (دانشیار گروه علوم اجتماعی دانشگاه اصفهان)

چکیده

این مقاله بر موضوع داده کاوی متمرکز است. داده کاوی فرآیند پیچیده جهت شناسایی الگوها و مدل‌های صحیح، جدید و به صورت بالقوه مفید، در حجم و سهولت از داده‌ها است. به طرفی که این الگوها و مدل‌ها برای انسان‌ها قابل درک باشند. داده کاوی، یک تکنیک از داده‌های هم‌کنش نامیده می‌شود. فرایند تحلیل داده‌ها از دیدگاه‌های متفاوت و خلاصه کردن آن‌ها به اطلاعات مناسب می‌باشد. اطلاعاتی که مورد انتقاد افراد گزارش و تقلیل هزینه ها مفید باشند. این مقاله بر اساس استاندارد (روش کتابخانه‌ای) و اطلاعات حاصل از مطالعات انجام شده انجام گرفت. بالاخره است، برای بسط پیشرفت از مطالعات هم از داده‌های موجود استفاده شد و پی در کل روش غالب در این مطالعه توصیفی و کتابخانه‌ای بتدریج. در مقاله حاضر پس از ارائه خلاصه آن از نتایج‌های میان روش‌های آماری و داده کاوی به زمینه‌های استفاده از داده کاوی، استاندارد کننده‌ی مراحل انجام داده کاوی و روش‌های استفاده از داده‌ها و همچنین محدودیت‌های آن اشاره شده است.

کلیدواژه‌ها: داده کاوی، کشف دانش، اطلاعات آماری، داده

مقدمه

در سال‌های اخیر در صنایع اطلاعاتی و جوامع مختلف داده کاوی به خاطر دسترسی کسترده به مقدار زیادی از داده‌ها و نیاز فوری به اطلاعات و شناخت مناسب بسیار مورد توجه قرار گرفته است که این شناخت و اطلاعات به دست آمده برای تحلیل‌های بزرگ و کاهش کلاهبرداری ها و ضبط و جذب مشتریان مفید بوده است(هان و کمبر، 2000: 5). در

1- Han & Kamber

مجله علوم اجتماعی دانشگاه آزاد، شماره 1341، صص 1-157
تاریخ دریافت: 1389/2/20
تاریخ تصویب: 1389/8/20
تاریخ نشر: 1390/6/20
طول دهه گذشته حجم وسیعی از داده‌ها در پایگاه‌های اطلاعاتی و داده‌ای جمع آوری شده‌اند که بیشتر آنها از ترم‌افزارهای تجاری، درخواست‌های مالی، مبنا و مدیریت تجاری و مدیریت روابط مشتری به دست آمده‌اند. نتیجه این جمع‌آوری داده‌های حجم‌اندازی این می‌تواند باشد که سازمان‌های و واحدهای تجاری، خدماتی و ... در این راستا اطلاعاتی غنی و لیل شناخت ضعیفی دارند و هدف اصلی داده‌کاوی این است که از روی داده‌های دم دست نسبت به استخراج اطلاعات در این مجموعه داده‌ها اقدام کرده و آنها را به شناخت تبیین نمایاند (ناگ و مک کلیتان، 2005: 2). دقت در استخراج داده‌ها جهت ترسیم و ضبط موجود و وضعیت اطلاعات با کمک افراد خبره و کاربران نهایی، موضوع مهمی است منجر به طرح روزی راهبردی به منظور حصول نتایج مورد نظر می‌باشد. با کمک داده کاوی بر روی پرستونه سازمان‌های مشابه و استخراج قواعد و دانش نهفته در آنها می‌توان به معاوضه در زمینه ارتباط بین بخش‌های مختلف در زمینه برنامه ریزی استراتژیک دست بafت. به عنوان یک روش برای انجام این کار می‌توان با توجه به معیار وضعیت فعلی با سه‌گزای مشکل‌ریزی را استخراج یا با بررسی با سه‌گزای افراد مربوط به نتایج دست یافته (سولیوان و دیگران، 1987: 75). داده کاوی یا کشف دانش در پایگاه داده‌ها ابزاری فنی و قدرتمند است برای استخراج دانش بالقوه‌های نهفته و اطلاعات پیشین سودمند از مجموعه‌ای از داده‌ها با کار می‌رود. این فرایند به صورت خودکار به یک روابط و اطلاعات موجود در داده‌های خام و اجرای نتایج آن می‌پردازد (فرانکلندز، 1952: 11). هدف داده‌کاوی شناخت ارتباطات و اطلاعات مکرر، تاریک، بالقوه‌سوزند و قابل فهم از داده‌های موجود می‌باشد. در اینجا عمومی کاوش داده‌ها به این کاذب گردیده حمل‌و نقل اطلاعات سازمان‌ها و مؤسسه‌ها و کار اشکال‌هایی که تعیین کردن درباره داده کاوی وجود ندارد. داده کاوی استخراج اطلاعات مفاهیم پنهان ناشناخته و به شکل بالقوه‌ست برای مجموعه‌ی از پایگاه‌ها به دست ها می‌باشد (لیو و چن، 2009: 373; زوانگ و همکاران، 2004: 225، جرفری). 3 665، که می‌تواند تحلیل‌های ارزشمندی را از میان حجم وسیعی

1- Tang and Macelmon
2- Fernandez
3- Lu & Chen
4- Zhuang et al
5- Jffrey
شاخص‌های اول

تکنیک داده‌کاوی و کاربرد آن در مطالعات اجتماعی

از داده‌ها به دست آورده (جراحی، ۱۹۵۲: ۲۲). به عبارت دیگر داده‌کاوی علم استخراج اطلاعات مفید از یافته‌های داده‌ای مجموعه‌ی داده‌ای می‌باشد (هنر و همکاران، ۲۰۰۱).

بنابراین داده‌کاوی، استخراج نیمه‌ای اطلاعاتی از داده‌ها. اگرایی، نامیده‌گری‌ها و دیگر ساختارهای معنی دار آماری از یافته‌های داده‌ها می‌باشد. بنابراین تعریف دیگر داده‌کاوی عملیات تجزیه و تحلیلی است که به منظور آشنایی با نکات کلیدی در یک موضوع کاری خاص و رطوحی شده اند.

توربان و همکارانش داده کاوی را فرایندی می‌دانند که در آن تکنیک‌های آماری، ریاضی، هرگیز مصنوعی و فناوری میانی برای استخراج و شناسایی اطلاعات مناسب و متعاقب کسب دانش از مجموعه‌ی داده‌ها به کار برده می‌شوند (ناوی، ۲۰۰۷: ۲۵۹۳). بریسون و همکارانش نیز تعریف مشابهی ارائه دادند. از نظر آنها واکاوی داده‌ها فرایند استخراج و آشنایی با الگوهای پنهان از مجموعه‌ی داده‌ها است (ناوی، ۲۰۰۹؛ ژیویون، ۲۰۰۱ و احمد، ۲۰۰۴). همچنین گونا گونه تکنیک‌های اثری کارش و جستجو در داده‌ها جهت استخراج دانش از آنها داده کاوی می‌گوید (گونا، ۲۰۰۶: ۴). هن و کمبر نیز داده کاوی را به استخراج و مuden کاری دانش از مقداری داده‌ها به تعریف کرده‌اند. بنابراین رابطه و دانش از داده‌های ناشناخته تعریف می‌گردد (هنر و کمبر، ۲۰۰۰: ۵). براساس تعریف مرکز تحقیقات آمریکا و اداره پاسخگویی سازمان‌ها داده کاوی به صورت دیگر تعریف می‌شود: داده کاوی مستلزم استفاده از ابزارهای پیشرفته تحلیل به کش روابط و الگوهای ارزش‌مند و ناشناخته در مجموعه‌ای داده‌ها است. بر این اساس داده‌کاوی نهایی شامل جمع آوری و مدیریت داده‌ها نیست، بلکه آن شامل تحلیل و پیش بینی داده‌ها نیز می‌باشد.

1- George
2- Hand et al
3- Ngai
4- Lejeune
5- Ahmed
6- Gupta
7- Han & Kamber
با توجه به تعاریف مطرح شده از دیدگاه‌های مختلف، می‌توان در جزء اساسی را در داده‌کاوی مشخص نمود. اولی کشف الگوهای پنهان در داده‌ها می‌باشد و دوم استفاده از این الگوها برای پیش‌بینی نتایج در آینده است.

در علوم اجتماعی داده‌ها اغلب حجم می‌باشد و به آسانی قابل استفاده نیستند. بلکه دانش‌نفیته‌ای که در داده‌ها وجود دارد، قابل استفاده می‌باشد. بنابراین بهتر گرایش از قدرت فرآیند داده کاوی جهت شناسایی الگوها و مدلهایی و نیز ارتباط عناصر مختلف در پایگاه داده جهت کشف دانش نفیته‌ای داده‌ها به دست آید و نهایتاً تبدیل داده به اطلاعات، روز به روز ضروری تر می‌شود. از آنجاییکه هر یک کوتاه‌ترین عناصر داده‌های داده‌کاوی است و با توجه به این که به کمک سیستم‌های کامپیوتری و پایگاه‌های داده، روزانه به میزان داده‌ها افزوده می‌شود، بنابراین استفاده هوشمندانه از دانش بالقوه ای که در این داده نفیته‌ای است در دنیای رقابی امروز شرکت‌ها حیاتی می‌باشد. داده‌های داده‌کاوی پیش‌بینی وضع آینده بازار، گرایش مشتریان و شناخت سلیقه‌های خریداری آنها را برای شرکت‌ها نمونه‌های خاص می‌سازد.

به طور کلی داده‌کاوی در تجارب به کار می‌رود و با استفاده از آن شرکت‌ها می‌توانند داده‌ها را بهتر کامپیوتر مورد الگوها و رفتار مشتریان را برای پیش‌بینی منابع استفاده، پیش‌بینی سود و جذب مشتری به کار گیرند (جورج، 1952: 2).
تشکیک داده کاوی و کاربرد آن در مطالعات اجتماعی

هدف اصلی در این مطالعه توصیف دقیق از داده کاوی و نقش و کاربرد آن در مطالعات علوم اجتماعی است. در این راستا اهداف فرعي تری مطرح است که در ذیل بیان می‌شوند:

- زمینه‌های به کارگیری داده کاوی در علوم اجتماعی کدامند؟
- مراحل اساسی انجام داده کاوی کدامند؟
- روش‌های منفی‌افته در داده کاوی کدامند؟

چرا از داده کاوی استفاده می‌کنیم؟

با توجه به این که در دهه‌های اخیر مؤسسات و سازمان‌های مختلفی توانسته‌اند مفاهیم و سیاست‌های جدیدی از داده‌های بازرگانی استفاده کنند. داده‌کاوی از آن‌ها یکی از بهترین ابزارهای فناوری‌های اجتماعی و به‌همین‌رو از دیدگاه های تاریخی-اجتماعی را به‌عنوان تکنولوژی‌های داده کاوی قابل توجهی در جهان امروزی دیده می‌شود.

در این راستا رمز مویقتی است که در این است که باید نگه داشتن مشتریان خود و هم‌چنین جذب مشتریان جدید از آن‌ها قوی داده کاوی استفاده کنند. داده‌کاوی از آن‌ها یکی از بهترین ابزارهای فناوری‌های اجتماعی و به‌همین‌رو از دیدگاه های تاریخی-اجتماعی را به‌عنوان تکنولوژی‌های داده کاوی قابل توجهی در جهان امروزی دیده می‌شود.

سازمان‌ها و مؤسسات خصوصی و دولتی به کار رود (تانگ و مک کلیمان، 2005: 4).

تفاوت بین داده و دانش

داده به هر آنچه که با مشاهدات جهان واقعی در ارتباط است، اطلاعات می‌شود و ماهیتی در حال تغییر و پویا دارد. کاملاً جزئی نگر است در حالیکه دانش به آنچه که کمتر دقیق است و پایداری بیشتری دارد. و در ارتباط با تعمیم ها، تجربه‌های داده است. نکته می‌شود؛ باعث حاصل از جدول 1 ارزش‌شناسی بودن دانش نسبت به داده را بیان می‌کند که ضرورت روش‌های داده کاوی را برای کشف دانش آشکار می‌سازد. طبق تعريف، به تکنیک‌های الیزخش کاوش و
جستجو در داده ها. جهت استخراج دانش از آنها، داده کاوی می گویند (رسولیان و دیگران، ١٣٨٧: ٢٨).

جدول ١: ویژگی‌های دانش و داده

<table>
<thead>
<tr>
<th>ویژگی‌ها</th>
<th>تفاوت</th>
</tr>
</thead>
<tbody>
<tr>
<td>به یک نمونه بر می‌گردد (یک نمونه از اشیاء، افراد و حوادث)</td>
<td>داده</td>
</tr>
<tr>
<td>ویژگی‌ها را جداگانه توصیف می‌کند.</td>
<td></td>
</tr>
<tr>
<td>با حجم زیاد در دسترس است (آرشیو و پایک اطلاعاتی).</td>
<td></td>
</tr>
<tr>
<td>امکان پیش بینی‌ها با ما نمی‌دهد.</td>
<td></td>
</tr>
<tr>
<td>به دسته‌ای مجموعه‌های نمونه بر می‌گردد.</td>
<td>دانش</td>
</tr>
<tr>
<td>ساختارها، اگرها و ... را به صورت عمومی توصیف می‌کند.</td>
<td>اغلب کسب یک مشکل است.</td>
</tr>
<tr>
<td>قابلیت پیش بینی دارد.</td>
<td></td>
</tr>
</tbody>
</table>

مأخوذ: (بیورن، ٢٠٠٢: ١٨)

پیشنهاد منهی

انتخاب اصطلاح ترکیبی تکنیک داده‌کاوی ناشی از تشابه‌ای وجود بین جستجوی اطلاعات ارزشمند در یک پایگاه داده بر یک و یا پایگاه داده بر یک و یا پایگاه داده بر یک مشابه است. هر دوی این کارها بر اینکه چه چیزی می‌تواند در مورد یا با چه نوعی کارکرد خلاقانه آنها برای تعیین دقیق محل قرار گرفتن چیزهای ارزشمند دلالت دارد (مدتی، ١٣٨٤: ٦١). یا یا در دیگر زمینه داده‌کاوی از اواخر دهه ١٩٩٠ شروع شد و پژوهش‌ها و مطالعات فراوانی در این راستا صورت گرفت و سیمینارها، دوره‌های آموزشی و کنفرانس‌های نیز برگزار شده است. استفاده از داده کاوی در سال‌های ٢٠٠٠ به بعد در مقایسه با آمار و در سال‌های ٢٠٠٢ به بعد به عنوان عاملی در رفتار مصرف کننده‌گان عرضه شد. در گذشته کشف دانش اغلب با استفاده از جمع جمعی اوری و ارزیابی داده‌ها آن‌ها در مقابل بسیاری پیش‌فرض‌های

1- Beveren
2- Data Mining
تعیین شده انجام می‌گرفته اما امروزه کشف دائمی با استفاده از روش‌های جدیدتری در مقابل رویکردی قدیمی مانند داده کاوی به عمل می‌آید (راجی و همکاران، 2002: 324). با این وجود جای این بحث همچنان در علوم اجتماعی خالی می‌باشد و در مطالعات اجتماعی مطالعات زیادی روی آن انجام نشده است.

زمینه‌هایی که کارگری داده کاوی

کاربرد داده کاوی اخیراً به‌طور وسیعی افزایش یافته است، در حالی‌که در ابتدا استفاده کنندگان از داده کاوی اساساً متعلق به صنایع اطلاع‌برداری بودند. مانند خدمات مالی و به‌صورت مستقیم در اختیار شرکت‌های بودا (جورج، 1952: 22). داده کاوی برای تنوع از اهداف در دوره‌های مختلف و عمومی به کار برده می‌شود. صنایعی چون بانکداری، بیمه، پزشکی و به طور عادی از داده کاوی برای کاهش هزینه‌ها و افزایش فروش استفاده می‌شود. برای مثال در بانکداری و بیمه از داده کاوی برای فاصله‌گذاری کلاه‌برداری و وضع کردن محدودیت‌های استفاده می‌شود (حرزی، 2004: 32). امروز، عملیات داده کاوی به صورت کلیه توسط تمامی شرکت‌های مالی، ارتباط‌های بانکداری و غیره استفاده می‌شود، از جمله فروشگاه‌ها. شرکت‌های مالی، ارتباط‌های بانکداری و غیره استفاده از داده‌های کاوی به‌این شرکت‌ها کمک می‌کند تا ارتباط عوامل داخلی و خارجی از جمله قیمت، محل فروشگاهی محصولات، هنر، کارمندان را با عوامل خارجی از جمله وضعیت اقتصادی، رقابت در بازار و محل جغرافیایی مشتریان کشف نمایند. هم‌چنین داده‌های کاوی با استفاده از تحقیقات مصرف کنندگان به‌گروه‌های با نیازهایی و ویژگی‌های مشتریان می‌توانند در بازاریابی خدمات ارزش‌آمیز ارائه دهند (نای و همکاران، 2009: 239).

یکی از نمونه‌های بارز داده کاوی را می‌توان فروش‌گاه‌های زنجیره‌ای مشاهده نمود، که در آن سعی می‌شود ارتباط محصولات مختلف هنگام خرید مشتریان مشخص گردد. فروش‌گاه‌های زنجیره‌ای مشتریان بدین‌گونه که چه محصولاتی با کیفیت‌گرده به فروش می‌روند. برای

1- Roger et al
مثال طی یک عملیات پایش داده‌های گسترده در یک فروشگاه زنجیره‌ای در آمریکای شمالی که بر روی حجم عمیقی از داده‌های فروش مجموعاً ایجاد خریداری می‌کند. همچنین مشخص کردن مشتری‌ای که تأثیر بیز بر ورود و عرضه پوشاک در اروپا مشاهده شده، به شکلی که نتایج داده‌های مشخص می‌کند که افرادی که کرایه‌های بریتیشی خریداری می‌کنند، در همان روز یا روزهای آینده گیره کرایات مشکی رنگ نیز خریداری می‌کنند. به روشنی این مطلب قابل درک است که این نوع استفاده از داده‌های فروش به کار خود فروشگاه‌ها را در برگزاری هوشمندانه فستیوال‌های فروش و نحوه ارائه اجناس به مشتریان پایدار رساند.

نمونه دیگر استفاده از داده‌های فروش در زمینه فروش را می‌توان در یک شرکت بزرگ دوبلاژ و تکنیک و عرضه فیلم‌های سینمایی در آمریکای شمالی مشاهده نمود که در آن عملیات داده کاوی، روابط مشتریان و هنری‌های سینمایی و نیز گروه‌های مختلف مشتریان بر اساس سیستم فیلم‌های (ترسکاک، رمانیک، هادلی ای و...) مشخص گردید. بنابراین آن شرکت به صورت کامل هوشمندانه می‌توانست مشتریان بالقوه فیلم‌های سینمایی را بر اساس علایق مشتریان به هنری‌های مختلف و سیستم سینمایی شناسایی کند. به طور کلی داده کاوی در زمینه‌های ذیل به کار می‌رود:

1- در زمینه‌های تجارت (بازاریابی هدف، تحلیل و مدیریت بازار، تحلیل سبد بازار، پیش بنی یکمیت نفت، فهم رفتار مشتری و تحلیل و مدیریت ریسک) با هدف کاهش هزینه پست‌‌ها موقتی نپایی گری‌هایی از مصرف کننده‌گان داده‌های کاوی به طور مومی می‌تواند در مورد اگرها و رفتار مشتریان و کاهش کلاه‌هایی، پیش‌بینی متغیب مورد استفاده و افزایش کسب سود مشتری و کنترل و ناپایه مشتریان به کار گرفته شود (فرنلاند، 1995: 11). به عنوان مثال کشف الگوی در خریده فروش‌ها برای شناسایی توی‌بات به ظاهر ناشتاخته که اگلب با هم خریداری شده‌اند، فهرست بندی کالاهای خریداری شده و هم چنین کالاهایی به فروش رسیده‌اند به این راستا مستلزم
استفاده از روش‌های داده‌کاوی است (همان: ۱۲). قابل ذکر است که خریدروشی‌ها نیز از داده‌کاوی برای تولیداتی به ظاهر ناشناخته که اغلب با هم خریداری شده‌اند و ارزیابی آن‌ها سه‌گاهی باعث افزایش آن‌ها که افزایش آن‌ها باعث افزایش آن‌ها
و ارتقاء (افزایش فروش تولیدات) استفاده می‌شود (جورج، ۱۹۵۲: ۲۴).

۲- ضمنه‌شناختی، مدیریت و کشف تخلف: (شناسایی تخلف‌های بی‌آموز فروش تولیدات و احتمالیت خریدن، احتمالیت خریدن، احتمالیت خریدن، برنامه‌ریزی، آزمون مشابه و کلیت که به نام‌هایی،..)

۳- در زمینه مستنداتی: (خلاصه‌سازی، یافتن نتایج مشابه و کلیت که به نام‌هایی،..)

۴- در زمینه‌های پزشکی: (کشف ارتباط و علائم پیماری، تحلیل آرایه‌های DNA و ساخت تصاویر پزشکی) داده‌کاوی می‌تواند پیش‌بینی کند که کمیابیک از مشتریان روش‌های جدیدی را خواهند خرید. لذا در روند رقابتی که مشتریان خطرنباید را مشخص می‌کند (فرانکنژر، ۱۹۵۳: ۱۳).

شهرت‌های داروسازی می‌تواند سوابق فروش اخیر خود را برای شناسایی دکترهای داروساز و تعیین این که گویانهای بازی‌های اثر را بر گری رشته‌ی تأیید کارشناسی باشد. افراد با خود به همین داده کاوی می‌تواند پیش‌بینی کند که کمیابیک از مشتریان روش‌های جدیدی را خواهند خرید (جورج، ۱۹۵۲: ۲۴).

۵- ارتباط و علائم بیماری، تحلیل آرایه‌های DNA و ساخت تصاویر بیماری (شکستگی) داده‌کاوی می‌تواند پیش‌بینی کند که کمیابیک از مشتریان روش‌های جدیدی را خواهند خرید. لذا در روند رقابتی که مشتریان خطرنباید را مشخص می‌کند (فرانکنژر، ۱۹۵۳: ۱۳).

شهرت‌های داروسازی می‌تواند سوابق فروش اخیر خود را برای شناسایی دکترهای داروساز و تعیین این که گویانهای بازی‌های اثر را بر گری رشته‌ی تأیید کارشناسی باشد. افراد با خود به همین داده کاوی می‌تواند پیش‌بینی کند که کمیابیک از مشتریان روش‌های جدیدی را خواهند خرید (جورج، ۱۹۵۲: ۲۴).

۶- تولیدات صنعتی کارخانه‌های کارخانه‌ها می‌تواند داده‌کاوی را برای بهبود فروش‌شان به خریدروش‌ها و ارائه یک محصول از طریق راه‌های مختلف می‌تواند شومه، راه آهن و پیش‌بینی کند (فرانکنژر، ۱۹۵۳: ۱۳).
خطر و سودجو بر اساس معیار هایی از جمله سن، درآمد، وضعیت سکونت، تحصیلات، شغل و غیره می‌انجامد.

tفاوت‌داده کاوش و تحلیل‌های آماری

داده کاوش توسط تجهیزات خاصی صورت می‌پذیرد که عملیات کاوش را بر اساس تجزیه و تحلیل مکرر داده‌ها انجام می‌دهد. داده کاوش با تحلیل‌های متداول آماری نیز متفاوت است. در تحلیل آماری، آمار شناسان همیشه با یک فرضیه شروع به کار می‌کنند. آنها از داده‌های علی‌اکثر استفاده می‌کنند و به دنبال پیدا کردن رابطه‌هایی هستند که به فرضیه آنها مربوط است. آنها می‌توانند داده‌های نابجا و نادرست را در طول تحلیل مشخص کنند و در نهایت نتایج کار خود را تفسیر و برای مدل‌سازی بیان کنند. اما در داده کاوش به فرضیه احتمالی نیست. ابزارهای داده کاوش از انتخاب مختلف داشت، یعنی، علی‌اکثر داده کاوش از انتخاب مختلف داشت، یعنی، علی‌اکثر داده کاوش استفاده کنند. گروه‌ها یا داده‌ها کاوش به طور اتوماتیک روابط را ایجاد می‌کنند. داده کاوش به داده‌های صحت و درست ناز دارد و این که نتایج داده کاوش‌ها نسبتاً پیچیده می‌باشد و نیاز به متخصصی جهت بیان آنها به مدیران دارد. داده کاوش اغلب برای ترکیبی از آمارهای (هوش مصنوعی و پژوهش‌هایی که در اینگونه انجام داده‌ها) مطرح شده بود. آن هم می‌توان این اشاره زمینه‌ای از علائم آماری شناخته و مطرح نشده بود و بعنوان اصطلاحی ناپویشانده در آمار مطرح شده بود.

(پیگیبون، 1997: 8).

به عنوان مثال در مورد شناخت کلاه‌داری‌های شرکت بیمه این دو روش نگاه متفاوتی دارند: در روش تحلیل آماری، یک محقق ممکن است موجه گردد که شرکت بیمه کلاه‌داری بیمه گردد. بر اساس این فرضیه، محقق به طرح یک سری سوال می‌پردازد تا این موضوع را بررسی کند. اگر نتایج حاصله مناسب بود، محقق فرضیه را اصلاح می‌کند و با اتخاذ فرضیه دیگری مجدداً شروع می‌کند. این روش به‌طور کامل وقت گیر است، بلکه به قدرت تجزیه و تحلیل محقق نیاز به‌طور کافی دارد. مهم‌تر از همه این که این روش هیچ وقت اگه‌های کلاه‌داری دیگری را که محقق به آنها مظنون نشده و در فرضیه جا نداده، بی‌پناه کند. اما در روش داده کاوش، یک

1- Pregibon
شماره اول

تحقیق سیستم‌های داده کاوی را ساخته و پس از طی مراحلی از جمله جمع‌آوری داده‌ها,

یکبارجه سازی و اخلاق داده‌ها به انجام عملیات داده کاوی می‌پردازد.

داده داده کاوی تمامگوهرای غیرعادی را که از حالت عادی و نرم‌ال انحراف دارند و ممکن

است منجر به کلبه‌داری شوند را پیدا می‌کند. نتایج داده کاوی حالت‌های مختلفی را که

تحقیق باید در مراحل بعدی پژوهش کند، نشان می‌دهند. در نهایت مدل‌هایی به دست آمده

می‌توانند مشتریانی را که امکان کلبه‌داری دارند، پیش بینی نمایند.

مثال دوم روش تحلیل سنتی: یک محقق می‌خواهد به مطالعه رفتار خرید یک طبقه

مشخص از مشتریان (مثلاً معلمان بازنشسته) برای طراحی بازار تبلیغاتی پردازد. ابتدا محقق از

خصوصیات شناخته شده این طبقه مشتری استفاده کرده و سعی می‌کند آنها را در گروه‌های

رده‌گذاری کند. سپس به بررسی رفتار خرید یکسان در هر یک از این گروه‌های پردازد. این کار

را آنقدر انجام می‌دهد تا به گروه‌های مناسب و مورد رضایتی بررسی کند.

اما ابزارهای داده کاوی به مطالعه بانک داده‌ها برای مشخص کردن تمام گروه‌هایی که

الگوی خرید مشخص دارند می‌پردازد. بعد از کاوش آن داده‌ها، محقق می‌تواند این نتایج را

گزارش دهد و با برای بررسی مجدداً به ابزارهای تحلیلی دیگری دهد.

نرم‌افزار داده‌های رفتاری از مهم ترین ابزارهای تحلیل برای داده‌ها است و به استفاده

کننده‌گان امکان می‌دهد که داده‌ها را از ابعاد و زوایای مختلفی تحلیل و دسته‌بندی و

روابط مشاهده شده بین آنها خلاصه کند.

مراحل اصلی انجام داده کاوی

داده‌های فرایندی تحلیلی است که برای کاوش داده‌ها (معمولاً حجم عظیمی از داده‌ها) در

زمینه‌های کسب و کار و بازار) صورت می‌گیرد و یافته‌ها با بکارگیری الگو‌هایی، ارزیابی اعتبار می‌شوند.

---

1 - Target Market
هدف اصلی داده کاوی پیش بینی است. داده کاوی را "کشف دانش در داده ها" نیز می‌نامند. کشف دانش داده ها در مراحل مختلفی می‌باشد که در اینجا به صورت خلاصه یا می‌سند:

1- شناخت مشکل و تعیین هدف موضوع: یکی از مهم ترین عواملی که باعث قصور داده‌ها می‌شود، عدم تعیین هدف بر اساس مشکلات کوتاه مدت و چندمدت شرکت هاست. داده کاوی به صورت اصلی از روش و شاخص هدف تجاری را تعیین می‌کند (فرنادز، 1952: 31). تا جایی که معلوم می‌کند تا چه اندازه شرکت‌ها انتظار متقایی دارد و داده کاوی چگونه به آنها کمک می‌کند.

2- یک سازی داده‌ها: کلید متقایی داده کاوی، استفاده از داده‌های مناسب است. در این مرحله داده‌های غیرمعنی‌دار از مجموعه داده‌های آموزشی خارج می‌شوند. به عبارتی این مرحله برای برطرف کردن خش و اطلاعات و داده‌های نامناسب به کار می‌رود (تالنگ و مک لینان، 2005: 7). داده‌هایی از اطلاعات ناکامی و نمونه‌هایی از داده‌های انتظاری که با پاکسازی در مورد آنها انجام گردید. معمولاً این مرحله با آماده سازی داده‌ها صورت می‌گیرد که ممکن است شامل پاس‌سازی داده‌ها، تبدیل داده واانتخاب زیرمجموعه‌هایی از رکوردها با حجم مطلوبی از متغیرها باشد (فرنادز، 1952: 31). سپس با توجه به ماهیت مسئله تحلیلی، این مرحله به مدل‌های پیش بینی ساده یا مدل‌های آماری و گرافیکی برای شناسایی متغیرها مورد نظر و تعیین پیچیدگی مدل‌ها برای استفاده در مرحله بعدی نیاز دارد.

3- انتخاب داده‌ها: این مرحله هدف کسب نمونه‌ای معرف با یکی از مناسب براز مدل‌سازی است. در این مرحله داده‌هایی که مربوط به فرآیند و اکتیو داده‌ها از مسیر داده‌ها جدا می‌شود. این مبحث را می‌توان بخشی از فرآیند کاهش اطلاعات نیز دانست.

4- تبدیل داده‌ها: داده‌ها به قابل قابل استفاده تبدیل می‌گردد. این مرحله صورت می‌گیرد، می‌توان به خلاصه سازی و یا محاسبه مقادیر تجمعی اشاره کرد. تبدیل داده‌ها، در این حالت متغیرهای پیوسته را در واحد‌های استاندارد شده بین می‌کند (فرنادز، 1952: 34).

1- Knowledge Discovery in data
2- Data Cleansing
3- Zoe et al
اتخاذ مدل: انتخاب مدل و الگوسازی هرنه اصلی در داده کاوار است. هند از مدل‌سازی در داده کاوار برای کشف روابط معنی‌دار عناصر داده‌ها است. این روابط که خصوصاً به الگوها و قوانین بر می‌گردد، می‌تواند برای پیش‌بینی و پیش‌بینی مودر استفاده باشد. (گزارش داده کاوار، 2008: 10). تشخیص الگوهای صحیح مورد نظر از سایر الگوها در این مرحله انجام می‌شود.

اصحاب الگوها بر اساس یک سری معیارهای جدایی‌سنجی می‌توانند (تاکوس و مک لیان، 2005: 16).

ارزیابی مدل: در نهایت قبل از این که مدل گسترش یابد و اجرا شود، باید ارزیابی شده و اعتبار آن سنجیده شود. در این مرحله با استفاده از محاسبات عدید و پارامترهای مختلف مدل‌هاي مختلفی به دست می‌آید. ارزیابی مدل به وسیله مجموعه‌ای از متغیرهای مستقل است که شرط لازم در داده کاوار برای تصمیم‌گیری و توان مدل می‌باشد (همان).

بازنمایی دانش: بازنمایی دانش را گزارش دهی نیز می‌نماید و آن یک نه برای نویسندگانش باشد.

در داده کاوار است. در این بخش به منظور ارائه دانش استخراج شده به کاربر، از یک سری ابزارهای بصری سازی استفاده می‌گردد (زوری و همکاران، 2009: 176). معمولاً در داده کاوار دو نوع گزارش داده می‌تواند گزارش دریافت‌یافته (الگوها) و گزارش هایی درباره پیش‌بینی یا پرواز داده‌های (تاکوس و مک لیان، 2005: 24).

کاهش اطلاعات در داده کاوار

کاهش اطلاعات عبارت است از تولید یک مجموعه کوچکتر از داده‌های اولیه که تحت عملیات داده کاوار نتایج تقریباً یکسانی با نتایج داده کاوار را اطلاعات اولیه به دست می‌دهد. این عمل را می‌توان از طریق حذف خصیصه‌های غیر مرتبط با نوع عملیات پایه داده‌های مورد نظر انجام داد. حذف خصیصه‌های مرتبط که در اثر انتخاب در ارزیابی میزان ارتباط آنها با عملیات داده کاوار انجام می‌گردد، می‌تواند منجر به ناکامی در انتخاب داده کاوار و استخراج قوانین ناقص و در نتیجه بر ارزش شود. عدم حذف خصیصه‌های مرتبط می‌تواند زمان انجام عملیات‌های داده کاوار را به طور قابل ملاحظه افزایش دهد. سه روش کلی برای انتخاب خصیصه‌های مرتبط با یکدیگر داده ها وجود دارد:

2
مجله علوم اجتماعی دانشکده ادبیات و علوم انسانی دانشگاه فردوسی مشهد
سال نهم

1. انتخاب بیش رونده: در هر مرحله خصیصه ای که بیشترین ارتباط را دارد، برگزیده می‌شود.
2. انتخاب عقب رونده: در هر مرحله خصیصه ای که کمترین ارتباط را دارد، انتخاب و حذف می‌شود.

3. روش ترکیبی: تركیب هر دو روش بیش و پس رونده

سلسله مراحل مفهومی نیز روشی برای کاهش تعداد مقادیر ممکنبرای یک خصیصه ارائه می‌دهد. اگر چه داده‌های خروجی کلی تربوه و فاقد برخی جزئیات هستند، اما این داده‌ها بسیار ساده تربوه و در سطح تجربی بالایی نسبت به داده‌های اولیه قرار دارند.

اطلاعات مورد نیاز برای عملیات داده کاری

1. داده‌های مرتبط با فرایند داده کاوی: با اطلاعات ممکن است شامل تعداد زیادی از رکوردها باشد که تنها یک خشک کوچکی از آنها با فرایند داده کاوی مرتبط هستند. مشخص کردن این بخش از اطلاعات باید توسط کاربر انجام گردد.
2. نوع دانش که باید استخراج شود: نوع روتین هایی که باید بر روی داده‌های انتخاب شده اعمال شوند. باید مشخص گردد.
3. دانش زمینه ای: کاربران می‌توانند با مشخص کردن دانش زمینه ای فرایند داده کاوی را هدایت نمایند. برای نمونه حاس کنار در مورد رفتار اطلاعات.
4. معیارهای ارزیابی دانش استخراج شده: این معیارها ممکن است در زمان اجرای فرایند واقعی داده‌های خروجی داده و باید ارزیابی شوند.
5. دانش را مشخص نمایند. دانش انتخاب شده، نمایش داده و قوانین استخراج شده در قلبی‌های مختلفی نظر جدول، نمودار، درخت تصمیم گیری و...

روش‌های مختلف داده کاری

این روشها بطور کلی به دو دسته زیر تقسیم می‌شوند:
شماره اول

تکیه‌گذاری و کاربرد آن در مطالعات اجتماعی

۱- الگوریتم‌های یادگیری با نظارت: در الگوریتم‌های یادگیری با نظارت، هدف از داده‌کاوب شرایط مشخص است و می‌خواهیم که به دنبال چه نوع داده‌کاوبی بررسی کنیم. روشهای یادگیری با نظارت شامل روش رگرسیون و طبقه‌بندی می‌باشد.

الف) مدلهای طبقه‌بندی: در مدلهای طبقه‌بندی می‌توان از احتمال عضویت طبقه‌ای به عنوان کلکسیون از متغیرهای درون‌داده مشخص نماید. این مدل‌ها برای پاسخ‌هایی که طبقه‌بندی شده‌اند مناسب هستند.

ب) رگرسیونی: در مدلهای رگرسیونی محقق می‌خواهد برآورد تقیی‌تر از تابع رگرسیونی داشته باشد. در حالیکه مدلهای رگرسیونی برای متغیرهای پیوسته و دوشابدی مناسب هستند، دو نوع از مدلهای رگرسیونی عبارتند از رگرسیون خطی و رگرسیون لجستیک. رگرسیون خطی متغیرهای پاسخ پیوسته را از روی تابع متغیرهای پیشین پیش بینی می‌کند. همچنین رگرسیون لجستیک را برای متغیرهای وابسته جفتی با ترتیب با کار می‌برد.

ج) مسیر تصمیم گیری و طبقه‌بندی که برای گسترش مسیر تصمیم گیری به شکاف‌های مجموعه‌ای کوچک‌تر از داده‌های کار می‌روند. هدف از مسیر تصمیم گیری مجموعه‌ای کوچک‌تر از داده‌های همگن با نوجوانی به متغیر هدف است. با این وجود، متغیرهای اساسی و طبقه‌بندی همانند متغیرهای پاسخ در مسیر تصمیم گیری به توانایی استفاده باشند.

(د) تحلیل داده‌کاوب‌های غیره: این نیز روشی طبقه‌بندی است که برای تبعیض متغیرهای پیشین در تمایز بین دو یا چند گروه پیشنهاد می‌کند. به‌عنوان مثال، مقدارهای مولکولی ای از اجزای دارند متغیر وابسته باشند و لیک نتایج مقدارهای طبقه‌ای و پیوسته می‌توانند به عنوان متغیر پیشین باشند.

و) آزمون کی: این یک روش طبقه‌بندی است که برای مطالعه رابطه بین سن‌سنجش پاسخ‌های مقدارهای مولکولی و متغیرهای پیشین احتمالی که ممکن است با هم ارتباط داشته باشند، مورد استفاده قرار می‌گیرد (فرنپادز، 1952: ۳۴).

از دیگر روشهای تحلیل در روش‌های یادگیری با نظارت در داده‌کاوب می‌توان به مدل‌سازی شبکه عصبی (که هم برای پیش بینی و هم برای طبقه‌بندی) ضرورت دارند. استفاده کردن.

1- Prediction Method
الف) تحلیل اجزاء اصلی: در تحلیل اجزاء اصلی ابعاد داده‌های چندمتغیری با استفاده از تبدیل
متغیرهای همبسته به متغیرهای ناهمبسته خطی کاهش می‌یابند.
ب) تحلیل عملي: در تحلیل عملي عامل‌های پنهان ناهمبسته کمی مقدار بیزی از واریانس‌های مشترک و جوابگو برای همبستگی مشاهده می‌شود داده‌های چندمتغیری که استخراج شده‌اند، به عهده می‌گیرند.
ج) تحلیل خوش‌آمیزی گسته‌تای: تحلیل های گسته‌تای ترکب موردها به گروه‌ها با خوش‌آمیزی
که هر گروه با خوش‌آمیزی هم در مشخصه‌های اصلی متناسب بوده، به کار می‌رود.
د) تحلیل های پیوستگی و سبد بازار: تحلیل سبد بازار یکی از راه‌بردها و مفاهیم اصلی تحلیل
های برای بازاریابی است. هدف تحلیل سبد بازار تعیین آن چه که تولید شده و خریداران نیز با هم
خرید می‌کنند، می‌باشند (فیلد: 1952: 73).

معرفی الگوریتم استخراج داده ها: داده‌های بدون پارامتر (پارامتر آزاد)
بسیاری از الگوریتم‌های داده کاری نیازمند تنظیم پارامترهای آزاد هستند. یک الگوریتم
بدون پارامتر یا پارامتر آزاد ما را از قرار گرفتن در معرض پیش داوری ناهنجار مسار در دست
برسی محفوظ می‌دارد و اجازه می‌دهد که خود داده ها ما صحبت کنند. این روش داده کاوبه
صورت تجسس حقيقی اجازه عمل می‌دهد تا این که ما را مجدی کند در معرض فرضیات داده‌ها
قرار بگیریم. یکی از مشکلات الگوریتم‌های پارامتر پنهان این است که تولید مجدد نتایج آزمایشی منتشر
شد است و این توزیع الگوریتم مشکل می‌باشد(کوک و همکاران، 2004: 22).

1- Description Methods
2- Keogh et al
شاخص اول

با وجود تعداد زیاد متغیرها، مشخص می‌توانید منفی‌نمای نتایج را درک کرد. بنابراین تحلیل گران از ایزاردی‌های نظیر کاهش ابعاد با تغییر شکل داده استفاده کرده و سعی می‌کند آنها را که تعدادشان تغییر نمی‌کند، نمایش دهد. این مراحل در مجموع اکتشاف دانش نامی‌دهد. اکتشاف دانش منظم عوامل مختلفی است از جمله:

1- همبستگی با یک فروردین وابستگی؟ که به آن تحلیل سبد بازار زیر می‌گویند. برای شناسایی مجموعه‌های مشترک از مقوله‌ها و قوانین برای هدف فروش به کار می‌روند(تانگ و مک لنان، 2005:4). هدف همبستگی برقراری روابط میان مقوله‌های که با هم‌دیگر در یک کارکرد مسلم فرض نشده است. تحلیل سبد خرید و برنامه‌های فروش مقطعی نمونه‌های بارزی برای مدل‌های همبستگی هستند(احمد، 2004). از این‌ها مورد استفاده در مدل‌های تحلیل همبستگی محاسباتی علت و معنایی آماری هستند.

2- طبقه‌بندی؟ یکی از مهم ترین مدل‌های معمول در داده‌های است. در این روش یک نمونه به یکی از چند دسته از پیش تعریف شده دسته بندی می‌شود. هدف این شاخه بیشینی رفتار آباده مصرف کننده یا طبقه بندی پایگاه‌های داده ای در شماری از طبقات از پیش تعریف شده براساس ضوابط معین است. مانند مدل چند روزه به چه رستوران های هومو و چه سفارشی می‌دهد. مهم ترین ایزاردی‌های که در این زمینه استفاده می‌شوند عبارتند از: شبکه‌های طبیعی، درخت تصمیم گیری و قانون "آکر/آن"، پس/آن"(ناتئ و همکاران، 2009:259 و تانک و مک لنان، 2000:2).

3- رگرسیون؟ پیش بینی یک مقدار متغیر مبنا بر متغیرهای دیگر. نوع تکنیک تخمین آماری است که برای هر موضوعی از داده‌ها ارزش‌های را پیش بینی می‌کند. موارد استفاده آن مانند متغیری بر۰، پیش بینی مدل سازی روابط علم و نسبت فرصتی علمی در مورد روابط بین متغیرها می‌باشد. از جمله می‌توان به رگرسیون خطی و رگرسیون لجستیک اشاره نمود.

1- Association Rule Discovery
2- Classification
3- Regression
4- خوشه بندي؛ که به اين قطعه قطعه سازي نيز ميگويد(همان، 4). موارد درون یک گروه در
آن یا-هم یکسان ولي با موارد خارج از آن گروه متفاوتند. خوشه بندي شامل فايند گروهيدي داده ها
به طبقات و خوشه هايي که موضوعات درون هر خوشه شاهب سپار بالاتري با همديديگر دندن یا با
دیگر خوشه ها متفاوت مي باشند(ژوي و همكاران، 2009. 373-383). خوشه ها گروه بنديهای دسته هاي
داده ای هستند که بر اساس شبات برخی از معيارها بوجود مي آيند. دسته بندي وظيفه قطعه قطعه
کردن گروهي ميگن به شماري از گروهي هاي همگن را به جهت دارد(کریس و پاول، 2003).
گروه بندي اسان ناشاخته پييين از جمله موارد خوشه بندي است(جرفي، 2004، 33). نفاوت اين
روش با طبقه بندي در اين است که در ابتدای محاسبات خوشه ها نتيم اتود و به عبارتی از پيي
تعريف شده ليستند. در این روش داده ها براساس روابط منطقي یا اولويت هاي مصير کننده تنظيمي
مي شوند. معرفت رست لينك ها در خوشه بندي شبکه هاي عصبی و تحليل مميز هستند(نائ و
همكاران، 2009).

5- تحليل دنباله؛ که به آن تحليل زنجيره ای نيز گفته مي شود. برایالگوهای موجود در یک
سری گستن به كار مي روند(ناگی و مک لینان، 2005، 5). ترتيب اتفاقاتي که در طول يك دوره
زمانی رخ مي دهد به عبارتی دیگر الگوهایي که در آن يك واقع خورد و فايند يكگرا به دنبال دارد
مانند تولد یک و خريدن گلف(جرفي، 2004، 23). اين روش الگوهای دنباله اي مانند سری های زمانی
را دنبال مي كند(ناگی و چن، 2009، 253-27). تحليل دنباله ها روابط يا الگوهای زمانی را نشان مي نشاند.
هدف آن بررسی مدل هاي است که در طول زمان ساخته مي شوند با برای استخراج و گزارش تغيير
مسیبها در طول زمان به كار مي رود. از جمله ابرهای مهم تحليل دنباله ها مي توان به تشويي
مجموعه ها اشاره نمود(ناگی و همكاران، 2009، 285). در اين روش داده ها براساس روندها و
الگوهای منتظم رفتار استخراج مي شوند. تفاوت ميان تحليل زنجيره اي و مدل هاي همبستگي در اين
است که مدل های زنجیره ای عناصر را جزي به جزي تحليل مي كنند اما مدل هاي همبستگي هر مقوله
را در يک چرخه خريد برای هم ارزی و مستقل بودن توضيح مي دهند(ناگی و مک لینان، 2005).

1- Clustering
2- Carrier & Povel
3- Sequential patterns
 مشکلات سیستم‌های داده کاوه

اگرچه واکاوی داده‌ها می‌تواند به روابط و گوهای آشکار کمک کند اما نمی‌تواند به استفاده کنندگان ارزش و اهمیت این گوهها را نشان دهد. به این دلیل گوههای کشف شده در این راستا وابسته به چگونگی آنها در مقایسه با شرایط جهان واقعی است. محدودیت دیگر در داده‌کاوهای این است که رابطه هایی را که بین متغیرها شناسایی کرده است، لزوماً رابطه‌ای علی‌به‌روی است. (جرفیر، 2004:۳۰) به طور کلی دو مشکل اصلی که اثر سیستم‌های داده‌کاوهی باید آن مواجه‌حساسیت، عبارتند از: یکی حجم بالای داده‌های آموزشی و پاک‌کردن اطلاعاتی به‌ورز و دوم وجود عدم قطعیت در اطلاعات. داده‌های عملیاتی موجود در سیستم‌های اطلاعاتی معمولاً دارای عدم قطعیت هستند. عدم قطعیت می‌تواند به اشکال مختلفی در پایگاه‌های داده ظهور کند. بطور کلی عدم قطعیت در سیستم‌های پایگاه داده به دو دسته تقسیم می‌شوند:

1- اطلاعات ناکام (مقاوم ناشی از): مظهر خصوصه‌هایی است که مقداری بی‌پایان آنها ثبت

نشده است.

2- اطلاعات ناسازگار: اطلاعاتی که در اثر اندوزه گیری نادرست یا بوجود آمدن نویز در داده‌ها

ایجاد شده باشد و مقدار ثابت شده با مقدار واقعی برابر نباشند.

برای رفع مشکلاتی که این سیستم‌ها در خروجی با داده‌های حجمی دارند، معمولاً روش‌های زیر استفاده می‌گردد:

1- طراحی گروپ‌های سریع: کاهش پجیک‌گی، بهینه‌سازی موازی سازی

2- کاهش حجم داده‌ها: نمونه‌گیری، کاهش سازی، کاهش ابعاد و ...

---

1- Jreffey
نتیجه‌گیری

روشنی این است که داده‌های هر یک از مراحل اولیه تکامل کودک به‌وجود می‌آمد و لذا موضع‌های مختلفی تحقیقاتی بسیاری را به وجود می‌آورد و هنوز کار زیادی از کار باید صورت پذیرد. همچنین داده کاری جوده‌ای مشکل در رشته‌های مختلف و زمینه تحقیقاتی عینی می‌باشد که به خصوص توجه گروه‌های تحقیقاتی سیستم‌های اطلاعات را به خود جلب کرده است. به علاوه، استفاده ورودی‌های به دست آمده توسط آمارگران که با داده‌ها سروکار دارند، نیاز به متخصصین کامپیوتر و پژوهش در این زمینه باشد. ارتباط با خبرگان سیستم‌های اطلاعاتی که به دلیل به کار آمدن با حدود انسبای و سیستم‌های موجود سخت است تنشگر این موضوع است. این سه مورد منجر به نتایجی می‌شود که به برخی محدودیت‌های مربوط به کارکرد داده‌ها اشاره دارند. نکته دیگر این که از داده کاوی به جای اکتشاف دانش از پانک اطلاعاتی هم استفاده می‌شود. این در زمانی است که از عبارت کارش داده‌ها برای روند عملیاتی در کسب و کار شرکت‌های جدید بهره می‌گیرند. همچنین هنگامی که اکتشاف دانش برای ایجاد تکنیک‌های جدید استفاده می‌شود، این عبارات مشترکا به جای هم به کار می‌روند.

منابع

رسولیان، محسن، ایوان‌نامه شرایع و محمدبافر فتحی گودری (1387): "نتیجه‌گیری داده‌های ابزارهای اکتشاف اندمیکی در مدیریت استراتژیک": فصل نامه عمومی، سال پانزدهم، شماره 39، صص. 4-101.

مقداسی، علی‌اکبر(1384): "نتیجه‌گیری داده‌های ابزارهای اکتشاف اندمیکی": Data Mining، صص. 51-65.


