فصلنامه مدیریت صنعتی دانشکده علوم انسانی، دانشگاه آزاد اسلامی
واحد سنندج – سال یازدهم / شماره ۳۷/ پاییز ۱۳۹۵

روش جدیدی برای رتبه‌بندی قواعد حاصل از داده‌کاوی با استفاده از تحلیل پوششی داده‌ها

حسین عزیزی
گروه ریاضی کاربردی، دانشگاه آزاد اسلامی، واحد پارس‌آباد مغان، پارس‌آباد مغان، ایران
Email: hazizi@iaupmogan.ac.ir

تاریخ دریافت: ۹۵/۴/۱۱ ❋ تاریخ پذیرش: ۹۵/۸/۱۹

چکیده

تکنیک‌های داده‌کاوی، یعنی استخراج الگوها از پایگاه‌های داده‌ای بزرگ، در تجارت به صورت گسترده‌ای مورد استفاده قرار می‌گیرند. با استفاده از این تکنیک‌ها ممکن است قواعد زیادی حاصل شوند و فقط تعداد کمی از آنها به دلیل محدودیت بودجه و منابع برای پیاده‌سازی در نظر گرفته شوند. ارزیابی و رتبه‌بندی جالب بودن و مفید بودن قواعد انجمنی در داده‌کاوی اهمیت زیادی دارد. در مطالعات قبلی که در مورد شناسایی قواعد انجمنی جالب از نظر ذهنی انجام شده است، اکثر روش‌ها مستلزم وارد کردن دستی یا پرسیدن از کاربر برای افتراق صریح قواعد جالب از ناجالب بوده است. این روش‌ها نیازمند محاسبات بسیار زیادی هستند و حتی ممکن است به نتیجه‌گیری‌های ناسازگار منتهی شوند. برای غلبه بر این مشکلات، این مقاله پیشنهاد می‌کند که از رویکرد تحلیل پوششی داده‌ها (DEA) با مرز دوگانه برای انتخاب کارآترین قاعده‌ی انجمنی استفاده شود. در این رویکرد علاوه بر بهترین کارآیی نسبی هر قاعده‌ی انجمنی، بدترین کارآیی نسبی آن نیز در نظر گرفته می‌شود. در مقایسه با DEAی سنتی، رویکرد DEA با مرز دوگانه می‌تواند کارآترین قاعده‌ی انجمنی را به درستی و به آسانی شناسایی کند. به عنوان یک مزیت، رویکرد پیشنهادی از نظر محاسباتی کارآمدتر از کارهای قبلی در این زمینه است. با استفاده از مثالی از تحلیل سبد بازار، قابلیت کاربرد روش مبتنی بر DEAی ما برای اندازه‌گیری کارآیی قواعد انجمنی با معیارهای چندگانه نشان داده خواهد شد.

کلمات کلیدی: تحلیل پوششی داده‌ها؛ DEA با مرز دوگانه؛ داده‌کاوی، قاعده‌ی انجمنی، جالب بودن، کارآیی‌های خوشبینانه و بدبینانه.

۱- مقدمه

با رشد سریع پایگاه‌های داده‌ای در بسیاری از تجارت‌های مدرن، داده‌کاوی تبدیل به روش مهمی در تحلیل داده‌ها شده است. در سال‌های اخیر، عرصه‌ی داده‌کاوی شاهد افزایش شدید علاقه هم از سوی دانشگاه و هم از طرف صنعت بوده است (Olafsson, Li, & Wu, 2008). افزایش حجم داده‌ها، افزایش آگاهی از ناتوانی ذهن انسان برای پردازش داده‌ها، و افزایش امکان یادگیری ماشینی، مهم‌ترین دلایل افزایش اقبال به داده‌کاوی بوده‌اند.

یکی از اهداف اصلی داده‌کاوی، ایجاد قواعد ارتباط از دیدگاه یک کاربر است. این کاربر بنا بر فرضیات موجود یک متخصص داده‌کاوی نیست، بلکه در رشته‌ای که تحت داده‌کاوی قرار دارد، تخصص دارد (Lenca, Meyer, Vaillant, & Lallich, 2008). مسئله‌ی پیدا کردن قواعد انجمنی تا حد زیادی مورد علاقه واقع شده و آلگوریتم‌های سریع متعددی برای داده‌کاوی قواعد انجمنی ابداع شده‌اند (Srikant, Vu, & Agrawal, 1997). با استفاده از این تکنیک‌ها می‌توان قواعد مختلفی به دست آورد و فقط تعداد کمی از این قواعد ممکن است برای پیاده‌سازی انتخاب شوند که این امر تا حدودی ناشی از محدودیت‌های بودجه و منابع است (M.-C. Chen, 2007). بنا به نوشته‌ی Liu و همکاران، مسئله‌ی جالب بودن از مدت‌ها به عنوان یک مسئله‌ی مهم در داده‌کاوی مطرح بوده است (B. Liu, Hsu, Chen, & Ma, 2000). معنای این مطلب، پیدا کردن قواعدی است که برای کاربر جالب/مفید باشند، نه اینکه صرفاً هر قاعده‌ای استخراج شود. در واقع، موقعیت‌هایی وجود دارد که به علت تعداد قواعد واجد شرایط (Tan, Kumar, & Srivastava, 2002) و محدودیت منابع (Choi, Ahn, & Kim, 2005)، لازم است که قواعد اولویت‌بندی شوند، تا قواعد مناسب انتخاب شوند و بر روی آنها تمرکز صورت گیرد. به نوشته‌ی Chen (۲۰۰۷)، انتخاب قواعد باارزش‌تر برای پیاده‌سازی، امکان موفقیت در داده‌کاوی را افزایش می‌دهد. به عنوان مثال، در تحلیل سبد بازار، فهمیدن اینکه مشتریان چه محصولاتی را با هم می‌خرند، و اینکه تبلیغات متقابل چه فوایدی برای فروشندگان دارد، می‌تواند تحلیلگران بازاریابی را به سوی خود جلب کند. روش اول به فروشندگان امکان می‌دهد که با در نظر گرفتن ترجیحات مشتریان، محصولات مناسب را به آنها عرضه کند، و روش دوم با در نظر گرفتن سودهای فروشندگان، امکان کسب نفع بیشتر را برای آنها ایجاد می‌کند. ترجیحات مشتریان را بر اساس پشتیبانی و اعتماد قواعد انجمنی می‌توان به دست آورد. از سوی دیگر، سود فروشندگان را می‌توان بر اساس اندازه‌های مربوط به دامنه از قبیل سود فروش و سود فروش متقابل مرتبط با قواعد انجمنی تعیین کرد (M.-C. Chen, 2007).

در مطالعات قبلی که در مورد شناسایی قواعد انجمنی جالب از نظر ذهنی انجام شده است، اکثر روش‌ها مستلزم وارد کردن دستی یا پرسیدن از کاربر برای افتراق صریح قواعد جالب از ناجالب بوده است (M.-C. Chen, 2007). Srikant و همکاران سه آلگوریتم یکپارچه را برای داده‌کاوی قواعد انجمنی با قیود مربوط به فقرات ارائه کردند (Srikant et al., 1997). به علاوه، Ng و همکاران رویکرد ارائه شده توسط Srikant و همکاران را بسط دادند تا امکان در نظر گرفتن قیود بسیار پیچیده‌تر، از جمله دامنه، دسته و قیود تجمیع به شیوه‌ی SQL، را داشته باشد (Ng, Lakshmanan, Han, & Pang, 1998 ; Srikant et al., 1997). Liu و همکاران یک سیستم تحلیل جالب بودن[1] را برای کمک به کاربر در شناسایی قواعد انجمنی جالب ارائه داده است (B. Liu et al., 2000). آنها در روش پیشنهادی خود، دو اندازه‌ی جالب بودن ذهنی اصلی، یعنی غیرمنتظره بودن و عمل‌پذیری، را در نظر گرفته‌اند. Choi و همکاران، با استفاده از فرآیند تحلیل سلسله‌مراتبی[2] روشی را برای اولویت‌بندی قواعد انجمنی ارائه کردند که ارزش‌های بازرگانی را در نظر می‌گیرد که متشکل از متریک عینی یا قضاوت‌های ذهنی مدیران هستند (Choi et al., 2005). آنها معتقد بودند که روش پیشنهادی آنها با تکنیک‌های تحلیل تصمیم برای حل مسئله در عرصه‌ی داده‌کاوی، هم‌افزایی می‌کند. با این وجود، این روش به مقدار زیادی تعامل انسانی برای به دست آوردن وزن معیارها با تجمیع نظارت مدیران مختلف، نیاز دارد.

روش‌های مختلف برای شناسایی کارآترین قاعده‌ی انجمنی مورد استفاده قرار گرفته است. یک راه، استفاده از تحلیل پوششی

[1] Interestingness analysis system (IAS).
[2] Analytic hierarchy process (AHP).

داده‌های[3] (DEA) سنتی برای شناسایی قواعد انجمنی کارآی خوشبینانه است، که بعد می‌توان کارآترین قاعده‌ی انجمنی را با استفاده از روش‌هایی مانند روش سوپرکارآیی شناسایی کرد. این روش باعث مشکلاتی مانند نشدنی بودن مدل‌ها می‌شود. همچنین، مستلزم آن است که تعداد زیادی مسئله‌ی برنامه‌ریزی خطی[4] (LP) حل شود. راه دیگر این است که از درخت‌های تصمیم، ماشین بردار پشتیبان، تکنیک‌های شبکه‌های عصبی مصنوعی و غیره برای شناسایی قاعده‌ی انجمنی استفاده شود. این موجب بروز مشکلاتی در شناسایی کارآترین قاعده‌ی انجمنی می‌شود. مشکل اصلی در استفاده از این روش‌ها، دشواری محاسباتی آنها است، زیرا به تعداد زیادی محاسبه‌ی وقت‌گیر نیاز دارند.

DEA یک تکنیک ناپارامتری مبتنی بر برنامه‌ریزی ریاضی برای اندازه‌گیری کارآیی نسبی مجموعه‌ای از واحدهای مشابه است، که معمولاً واحدهای تصمیم‌گیری[5] (DMUها) نامیده می‌شوند. DEA به خاطر کاربردها و مطالعات موردی موفق خود، از سوی محققان صنعتی و دانشگاهی مورد توجه بسیار زیادی قرار گرفته است. ارزیابی عملیات انبارداری داده‌ها (Mannino, Hong, & Choi, 2008)، انتخاب سیستم تولید انعطاف‌پذیر (S.-T. Liu, 2008)، سنجش عملکرد شعب بانک (Camanho & Dyson, 2005)، بررسی کارآیی بانک (X. Amirteimoori, Kordrostami, & Azizi, 2016 ; Chen, Skully, & Brown, 2005)، تحلیل اظهارنامه‌های مالی بنگاه‌ها (Edirisinghe & Zhang, 2007)، اندازه‌گیری کارآیی مؤسسات آموزش عالی (Johnes, Amirteimoori, Emrouznejad, & Khoshandam, 2013 ; 2006)، حل مسئله‌ی طراحی چینش تأسیسات (Ertay, Ruan, & Tuzkaya, 2006)، و اندازه‌گیری کارآیی سرمایه‌گذاری‌های سازمانی در فناوری اطلاعات (Shafer & Byrd, 2000) نمونه‌هایی از کاربرد DEA در عرصه‌های مختلف هستند. Chen (۲۰۰۷) کار Choi و همکاران (۲۰۰۵) را توسعه داد و روشی را بر پایه‌ی DEA برای رتبه‌بندی قواعد انجمنی با در نظر گرفتن معیارهای متعدد ارائه کرد. وی در طی روش رتبه‌بندی خود از یک مدل DEAی پیشنهاد شده توسط Cook و Kress برای شناسایی قواعد انجمنی کارآی خوشبینانه استفاده می‌کند (Cook & Kress, 1990). آنگاه یک مدل DEAی دیگر را که به وسیله‌ی Obata و Ishii ابداع شده است، اعمال می‌کند، تا بین قواعد انجمنی کارآی خوشبینانه افتراق قایل شود (Obata & Ishii, 2003). شایان توجه است که روش پیشنهادی او مستلزم آن است که مدل اول برای همه‌ی DMUها حل شود و مدل دوم برای DMUهای کارآی خوشبینانه حل شود. عیب این روش آن است که نیاز به حل تعداد زیادی مسئله‌ی LP دارد. به علاوه، این رویکرد شامل مقداری محاسبات و ملاحظات زاید است. این مقاله نیز، مانند Chen (۲۰۰۷)، از DEA به عنوان یک رویکرد پس‌پردازش استفاده می‌کند. پس از آنکه قواعد از آلگوریتم‌های داده‌کاوی استخراج شدند، از DEA برای رتبه‌بندی آنها بر اساس برخی معیارهای تعیین شده استفاده می‌شود.

برای اینکه از DEA بهترین بهره گرفته شود، ما در این مقاله روش DEAی جدیدی را که "DEA با مرز دوگانه" نامیده می‌شود، برای شناسایی کارآترین قاعده‌ی انجمنی معرفی می‌کنیم. DEA با مرز دوگانه دو کارآیی را برای تصمیم‌گیری در نظر می‌گیرد: یکی نسبت به مرز تولید کارآ اندازه‌گیری می‌شود و *بهترین کارآیی نسبی* یا *کارآیی خوشبینانه* نامیده می‌شود، و دیگری نسبت به مرز تولید ناکارآ، که به آن مرز ورودی نیز می‌گویند، سنجیده می‌شود و *بدترین کارآیی نسبی* یا *کارآیی بدبینانه* نامیده می‌شود. DEAی سنتی فقط بهترین کارآیی‌های نسبی گروهی از DMUها را ضمن اجتناب از کارآیی‌های بدبینانه اندازه‌گیری می‌کند، بنابراین، نمی‌تواند یک سنجش کلی از DMUها ارائه دهد. با در نظر گرفتن همزمان کارآیی‌های خوشبینانه و بدبینانه، همه‌ی DMUها را می‌توان بدون نیاز به محاسبات زیاد، به طور کامل رتبه‌بندی کرد. این را با مثال عددی نشان خواهیم داد.

داده‌کاوی قاعده‌ی انجمنی که به وسیله‌ی Agrawal، Imielinski و Swami، ارائه شده است (Agrawal, Imieli, & Swami, 1993)، در کاربردهای تجاری سنتی، از قبیل بازاریابی متقابل، نامه‌رسانی متصل، طراحی کاتالوگ، تحلیل رهبر زیان، چینش فروشگاه، و تقسیم‌بندی مشتریان برای کاربردهای بازرگانی مانند تجدید صفحات وب و شخصی کردن وب مورد استفاده قرار گرفته است (Choi et al., 2005).

---

[3] Data envelopment analysis (DEA).
[4] Linear programming (LP).
[5] Decision-making units (DMUs).

با داشتن مجموعه‌ای از تراکنش‌ها، که هر تراکنش مجموعه‌ای از الفاظ (به نام فقره) است، قاعده‌ی انجمنی بیانی به صورت $X \Rightarrow Y$ است، که در آن X و Y مجموعه‌های فقرات هستند. معنای شهودی چنین قاعده‌ای آن است که تراکنش‌های پایگاه داده‌ای که حاوی X هستند، حاوی Y هستند. نمونه‌ای از قاعده‌ی انجمنی به این صورت است: «۴۰٪ تراکنش‌هایی که حاوی نان هستند، حاوی شیر هم هستند؛ ۳٪ کاهش تراکنش‌ها حاوی هر دو فقره هستند.» در اینجا، ۴۰٪ اعتماد قاعده نامیده می‌شود، و ۳٪ پشتیبانی قاعده نام دارد. باید در نظر داشت که قواعد انجمنی هر تعداد فقره در هر طرف ممکن است داشته باشند. یک آلگوریتم مؤثر مورد نیاز است که فضای جستجو را محدود کند و فقط زیرمجموعه‌ای از همه‌ی قواعد انجمنی را کنترل کند، و در عین حال، هیچ قاعده‌ی انجمنی مهمی را از دست ندهد (M.-C. Chen, 2007). برای کشف قواعد انجمنی از داده‌ها و به دست آوردن الگوهای مفید از آلگوریتم‌های مختلفی می‌توان استفاده کرد. آلگوریتم Apriori یکی از متداول‌ترین و مشهورترین تکنیک‌ها برای پیدا کردن قواعد انجمنی است (Agrawal & Srikant, 1994Agrawal et al., 1993 ;). این روش در دو مرحله عمل می‌کند: در مرحله‌ی اول، تمام مجموعه‌های فقرات با مینیموم پشتیبانی (مجموعه‌های فقرات مکرر) تولید می‌شوند. در این مرحله از خاصیت بسته شدن پایین پشتیبانی استفاده می‌شود. به عبارت دیگر، اگر مجموعه‌ی فقرات با اندازه‌ی k یک مجموعه‌ی فقرات مکرر باشد، آنگاه همه‌ی مجموعه‌های فقرات با اندازه‌ی زیر $(k-1)$ نیز باید مجموعه‌های فقرات مکرر باشند. با استفاده از این خاصیت، مجموعه‌های فقرات نامزد با اندازه‌ی k از مجموعه‌ی فقرات مکرر با اندازه‌ی $(k-1)$ با تحمیل این شرط که همه‌ی زیرمجموعه‌های فقرات با اندازه‌ی $(k-1)$ از هر مجموعه‌ی فقرات نامزد باید در مجموعه‌های فقرات مکرر با اندازه‌ی $(k-1)$ وجود داشته باشند، ساخته می‌شود.

داده‌کاوی قاعده‌ی انجمنی تکنیک متداولی برای تحلیل سبد بازار است، که عموماً هدف آن پیدا کردن الگوهای خرید از سوپرمارکت‌ها، سفارش پستی و مشتریان دیگر است. با داده‌کاوی قواعد انجمنی، تحلیلگران بازار سعی می‌کنند مجموعه‌ای از محصولات را که غالباً با هم خریده می‌شوند، شناسایی کنند، تا بر اساس اقلام موجود در سبد خرید، بتوان فقرات مناسب دیگر را استنباط کرد. مثلاً قواعد انجمنی را غالباً می‌توان در تبلیغات بازاریابی با چینش مناسب محصولات روی قفسه‌ی سوپرمارکت و هدایت کردن اقلام پیشنهادی به مشتریان که ممکن است مورد علاقه‌ی آنها باشد، مورد استفاده قرار داد (M.-C. Chen, 2007).

داده‌کاوی در تجارت الکترونیک: داده‌کاوی به شرکت‌ها امکان می‌دهد که الگوهای نهفته در درون تراکنش‌های خرید گذشته را بفهمند، و از این طریق، به برنامه‌ریزی و راه‌اندازی پویش‌های بازاریابی جدید به صورت عاجل و مقرون به صرفه کمک می‌کند (Padhy, Mishra, & Panigrahi, 2012). تجارت الکترونیک یکی از آینده‌نگرترین حوزه‌ها برای داده‌کاوی است، زیرا در این عرصه، سوابق داده‌ها، از جمله داده‌های مشتریان، داده‌های محصول، داده‌های لاگ اقدامات کاربران، فراوان‌اند؛ تیم فناوری اطلاعات مهارت غنی داده‌کاوی دارد و می‌توان بازگشت سرمایه را اندازه‌گیری کرد. پژوهشگران از تحلیل انجمنی[۶] و خوشه‌بندی[۷] بهره می‌گیرند تا مشخص کنند چه محصولاتی در ترکیب با هم خریداری شده‌اند؛ با این کار، می‌توان مشتریان را تشویق کرد که محصولات مرتبط را که ممکن است از دیدشان پنهان مانده باشد، خریداری کنند. رفتار کاربران پایش و تحلیل می‌شود، تا شباهت‌ها و الگوها در رفتار وِب‌نوردی آنها شناسایی شود، به طوری که بتوان وب را در تأمین نیازهای کاربران موفق‌تر کرد (Heer & Chi, 2001). یک روش مکمل، شناسایی محتوایی است که می‌تواند برای کاربر جالب باشد، و در آن از داده‌های مربوط به ترجیحات مجموعه‌ای از کاربران استفاده می‌شود، و به آن پالایش گروهی[۸] یا سامانه‌های توصیه‌گر[۹] می‌گویند (Breese, Heckerman, & Kadie, 1998; Resnick & Archak, Ghose, & Ipeirotis, 2011 ; Varian, 1997). این روش با استفاده از همبستگی و دیگر سنجه‌های مشابهت کاربران، پروفایل‌های کاربران مشابه را شناسایی و خوشه‌بندی می‌کند تا از آن برای توصیه‌ی اقلام اطلاعاتی به کاربران استفاده شود. و سیستم توصیه‌گر به شبکه‌ی

[6] Association analysis
[7] Clustering
[8] Collaborative filtering
[9] Recommender systems

اجتماعی (Guy, 2014)، عرصه‌ی آموزش (Konstan, Walker, Brooks, Brown, & Ekstrand, 2014)، کتابخانه‌های دانشگاهی (Tejeda-Lorente, Bernabé-Moreno, Porcel, & Herrera-Viedma, 2014)، و گردشگری (Gavalas, Konstantopoulos, Mastakas, & Pantziou, 2014) نیز بسط داده می‌شود.

داده‌کاوی در صنعت: داده‌کاوی می‌تواند برای صنایعی مانند فروشگاه‌ها، بانکداری، و مخابرات بسیار سودمند باشد: طبقه‌بندی و خوشه‌بندی را می‌توان در این عرصه نیز به کار گرفت (Elgendy & Elragal, 2014).

یکی از عناصر کلیدی موفقیت سازمان‌های بیمه و بانک‌ها، سنجش قبلیِ اعتبار وام گیرندگان در طی فرآیند ارزیابی اعتبار است. نمره‌دهی اعتبار اهمیت بیشتر و بیشتری پیدا می‌کند و چندین روش داده‌کاوی برای مسئله‌ی نمره‌دهی اعتبار مورد استفاده قرار می‌گیرد (Kambal, Osman, Taha, Mohammed, & Mohammed, 2013; Koh, Hsieh & Hung, 2010 ; Tan, & Goh, 2006).

فروشگاه‌داران از اطلاعات مشتریان، اطلاعات تراکنش‌های مرتبط، و اطلاعات محصول استفاده می‌کنند تا دقت پیش‌بینی تقاضا برای محصول، بهینه‌سازی دسته‌های کالاها، توصیه‌ی محصول، و رتبه‌بندی فروشگاه‌ها و تولید کنندگان را به طور قابل توجهی بهبود بخشند (Maaß, Spruit, & de Waal, 2014Q. Liu, Wan, & Zhou, 2014 ;). پژوهشگران از SVM (Du, Leung, Zhang, & Lai, 2013)، رگرسیون بردار پشتیبانی (Lu & Wang, 2010)، یا مدل بیس[10] (Lee, Kim, Park, & Kang, 2014) برای پیش‌بینی تقاضا برای محصول استفاده می‌کنند.

داده‌کاوی در مراقبت بهداشتی: در عرصه‌ی بهداشت و درمان، داده‌کاوی به طور روزافزونی مورد اقبال واقع می‌شود و چه بسا تبدیل به یکی از عناصر ضروری شده است (M. Chen, Gonzalez, Leung, Zhang, & Li, M. Chen, 2014 ; 2010; M. Chen, Ma, Jialun, Dung Ong, & Song, 2013; M. Chen, Mau, Wang, & Wang, 2013; J. Liu, Wan, He, & Zhang, 2014; J. Liu, Wang, Wan, Xiong, & Zeng, 2013). داده‌های ناهمگن پزشکی در سازمان‌های مختلف بهداشتی و درمانی ایجاد شده‌اند، از جمله اطلاعات پرداخت کنندگان، کارشناسان پزشکی، اطلاعات داروسازی، اطلاعات تجویز داروها، یادداشت‌های پزشکان، و یا پرونده‌های بالینی که به صورت روزمره ایجاد می‌شوند. از این داده‌های کمّی می‌توان برای متن‌کاوی بالینی، مدل‌سازی پیش‌بینانه (Wan et al., 2013)، تحلیل بقا، تحلیل شباهت بیماران (Duan, Street, & Xu, 2011)، و کاهش زباله استفاده کرد. در عرصه‌ی مراقبت بهداشتی، تحلیل انجمنی، خوشه‌بندی، و تحلیل برون‌هشته‌ها را می‌توان مورد استفاده قرار داد (Kincade, 1998 ; Sun & Reddy, 2013;).

داده‌های پرونده‌های درمان را می‌توان مورد داده‌کاوی قرار داد، تا روش‌هایی برای کاهش هزینه‌ها و ارائه‌ی درمان بهتر شناسایی شود (Jun Liu et al., 2013Bellazzi & Zupan, 2008 ;). همچنین، می‌توان از داده‌کاوی برای شناسایی و درک بیماران پرهزینه استفاده کرد (Silver et al., 2001)، و آن را بر روی انبوه داده‌های حاصل از میلیون‌ها نسخه، عمل جراحی، و دوره‌های درمانی اعمال کرد تا بتوان الگوهای غیرمعمول و کلاهبرداری‌ها را شناسایی کرد (Koh & Tan, 2011 ; Thornton, Mueller, Schoutsen, & van Hillegersberg, 2013).

داده‌کاوی در نظارت شهری: در عرصه‌ی خدمات عمومی، از داده‌کاوی می‌توان برای کشف نیازهای عمومی و بهبود عملکرد خدمات، تصمیم‌گیری با سیستم‌های خودکار برای کاهش خطرات، طبقه‌بندی، خوشه‌بندی، و تحلیل سری‌های زمانی به منظور حل مسایل در این عرصه استفاده کرد.

دولت الکترونیک کیفیت نظارت بر خدمات، صرفه‌جویی هزینه‌ها، مشارکت سیاسی بیشتر، و سیاست‌ها و برنامه‌های کارآمدتر را بهبود می‌بخشد (Wan, Li, Zou, & Zhou, 2012Helbig, Ramón Gil-García, & Ferro, 2009 ;)، و همچنین، به عنوان روشی برای افزایش ارتباط شهروندان با نهادهای دولتی و نهایتاً اعتماد سیاسی پیشنهاد شده است (Chadwick & May, 2003). سیستم مدیریت اطلاعات حوادث شهری می‌تواند با روش‌های داده‌کاوی تلفیق شود تا سنجش جامعی از تأثیر بلایای طبیعی بر تولید کشاورزی ارائه کرده و نواحی مصیبت‌زده را به صورت عینی رتبه‌بندی کند و به

[10] Bass

دولت‌ها به منظور آمادگی حوادث و تخصیص منابع کمک کند (Peng, Zhang, Tang, & Li, 2011).
پژوهشگران با استفاده از تحلیل داده‌ها می‌توانند پیش‌بینی کنند که کدامیک از شهروندان ممکن است از شهر خارج شوند (Sullivan & Mitra, 2014)، و بر این اساس، می‌توان استنباط کرد که چه عواملی از زندگی و خدمات شهری منجر به تصمیم شهروندان برای خروج از شهر می‌شود (M. Chen, 2013).
یک چالش بزرگ برای دولت و نیروی انتظامی، چگونگی تحلیل سریع حجم رو به افزایش داده‌های مربوط به جرایم است (H. Chen et al., 2004). پژوهشگران از تکنیک‌های داده‌کاوی فضایی برای پیدا کردن قواعد ارتباط بین نقاط داغ جرم و جنایت و چشم‌انداز فضایی استفاده می‌کنند (Huang, 2013)؛ برخی دیگر از پژوهشگران از آلگوریتم تقویت شده‌ی خوشه‌بندی میانگین‌های k برای کشف الگوهای جرم و استفاده از تکنیک یادگیری نیمه‌نظارت‌شده برای کشف دانش و کمک به افزایش دقت پیش‌بینی استفاده می‌کنند (Shyam, 2006). همچنین، از داده‌کاوی می‌توان از طریق تحلیل اطلاعات افراد، از قبیل نام، نشانی، تاریخ تولد، و کد ملی، برای شناسایی موارد کلاهبرداری مجرمانه‌ی هویت (G. Wang, Chen, & Atabakhsh, 2004) و شناسایی الگوهای ساختاری ناشناخته‌ی شبکه‌های جنایی استفاده کرد (Hsinchun Chen et al., 2003).
در سیستم حمل و نقل، از داده‌کاوی می‌توان برای اصلاح نقشه بر اساس ردیابی‌های GPS استفاده کرد (Cao, Cong, & Schroedl, Wagstaff, Rogers, Langley, & Wilson, 2004; Wan, Zhang, Zhao, Jensen, 2010 ; Yang, & Lloret, 2014)، و پژوهشگران بر اساس مسیرهای GPS کاربران، محل‌های جالب و توالی‌های کلاسیک مسافرت را به منظور توصیه‌ی محل و توصیه‌ی سفر شناسایی می‌کنند (Zheng, Zhang, Wie, & Ma, 2009).
کاربرد داده‌کاوی و تکنیک‌های متداول داده‌کاوی در جدول ۱ به طور خلاصه بیان شده است.

جدول شماره(۱): کاربرد داده‌کاوی و تکنیک‌های متداول داده‌کاوی.

کاربرد	طبقه‌بندی	خوشه‌بندی	تحلیل انجمنی	تحلیل سری زمانی	تحلیل برون‌هشته‌ها
تجارت الکترونیک		✓	✓		
صنعت	✓	✓	✓		
مراقبت بهداشتی		✓	✓		✓
نظارت شهری	✓	✓	✓	✓	

۲- مواد و روش‌ها

DEA یک رویکرد مبتنی بر داده‌ها برای ارزیابی نسبی عملکرد گروهی از موجودیت‌ها یا اصطلاحاً DMUها است. این روش بر اساس کارهای پیشگامانه‌ی Farrell به وسیله‌ی Charnes، Cooper و Rhodes ابداع شد (Charnes, Cooper, & Farrell, 1957Rhodes, 1978 ;). آنها تعریف نسبت خروجی منفرد به ورودی منفرد را که برای کارآیی ارائه می‌شد، برای حالت دارای ورودی‌ها و خروجی‌های متعدد تعمیم دادند. Charnes و همکاران در مدل DEAی اولیه‌ی خود (مدل CCR) پیشنهاد کردند که کارآیی یک DMU را می‌توان به صورت ماکزیموم نسبت خروجی‌های وزن داده شده به ورودی‌های وزن داده شده به دست آورد، مشروط بر آنکه آن نسبت برای همه‌ی DMUها باید کمتر یا مساوی یک باشد (Charnes et al., 1978). پوشش در CCR به صورت بازده به مقیاس ثابت است، یعنی یک افزایش متناسب در ورودی‌ها منجر به افزایش متناسبی در خروجی‌ها می‌شود. Banker، Charnes و Cooper مدل BCC را برای ارزیابی کارآیی فنی خالص DMUها با اشاره به مرز کاراً ابداع کردند (Banker, Charnes, & Cooper, 1984). به علاوه، این مدل مشخص می‌کند که یک DMU در شرایط بازده به مقیاس افزایشی، کاهشی یا ثابت عمل می‌کند. بنابراین، مدل‌های CCR نوع خاصی از مدل‌های BCC هستند.
فرض کنید n DMU برای انتخاب وجود دارند، که باید از نظر m ورودی و s خروجی ارزیابی شوند. برای DMU_j ($j=1,\ldots,n$)، مقادیر ورودی را با نماد x_{ij} ($i=1,\ldots,m$) و مقادیر خروجی را با نماد y_{rj} ($r=1,\ldots,s$) نشان می‌دهیم،

که همه‌ی آنها معلوم و نامنفی هستند. کارآیی DMU_j به صورت زیر تعریف می‌شود:

$$\theta_j = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \qquad (۱)$$

که در اینجا u_r ($r=1,\ldots,s$) و v_i ($i=1,\ldots,m$) وزن‌هایی هستند که به ترتیب برای s خروجی و m ورودی تعیین شده‌اند.

کارآیی خوشبینانه‌ی DMU_j نسبت به DMUهای دیگر با مدل CCR زیر سنجیده می‌شود (Charnes et al., 1978):

$$\begin{aligned} \max \quad & \theta_o = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \\ \text{s.t.} \quad & \theta_j = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \quad j = 1,\text{K}\ ,n, \\ & u_r, v_i \ge \varepsilon, \quad r = 1,\text{K}\ ,s; \quad i = 1,\text{K}\ ,m, \end{aligned} \qquad (۲)$$

که در اینجا DMU_o به DMU تحت ارزیابی اشاره دارد، و u_r و v_i متغیرهای تصمیم هستند. همچنین، ε یک مقدار بی‌نهایت‌کوچک غیرارشمیدسی برای جلوگیری از صفر شدن وزن‌ها است. با استفاده از تبدیل Charnes و Cooper، مدل برنامه‌ی کسری (۲) را می‌توان به LP زیر تبدیل کرد (Charnes & Cooper, 1962):

$$\begin{aligned} \max \quad & \theta_o = \sum_{r=1}^{s} u_r y_{ro} \\ \text{s.t.} \quad & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1,\text{K}\ ,n, \\ & \sum_{i=1}^{m} v_i x_{io} = 1, \\ & u_r, v_i \ge \varepsilon, \quad r = 1,\text{K}\ ,s; \quad i = 1,\text{K}\ ,m. \end{aligned} \qquad (۳)$$

اگر مجموعه‌ای از وزن‌های مثبت u_r^* ($r=1,\ldots,s$) و v_i^* ($i=1,\ldots,m$) وجود داشته باشند تا $\theta_o^* = 1$ را تأمین کنند، آنگاه DMU_o کارآی DEA یا کارآی خوشبینانه نامیده می‌شود؛ در غیر این صورت، به آن غیرکارآی DEA یا غیرکارآی خوشبینانه می‌گویند. کارآیی خوشبینانه، در مقالات DEA، کارآیی CCR نیز نامیده می‌شود. برای n DMU مختلف، مدل LP (۳) جمعاً n بار، هر بار برای یک DMU، حل می‌شود. همه‌ی DMUهای کارآی خوشبینانه یک مرز تولید کارآ را تشکیل می‌دهند، و می‌توان آنها را برای تحلیل بیشتر با هم گروه‌بندی کرد.

کارآیی بدبینانه‌ی DMU_j نسبت به DMUهای دیگر با مدل کارآیی بدبینانه‌ی زیر اندازه‌گیری می‌شود (Amirteimoori, 2007; Azizi, 2011; Y.-M. Wang & Yang, 2007):

$$\begin{aligned} \min \quad & \varphi_o = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \\ \text{s.t.} \quad & \varphi_j = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \ge 1, \quad j = 1,\text{K}\ ,n, \\ & u_r, v_i \ge \varepsilon, \quad r = 1,\text{K}\ ,s; \quad i = 1,\text{K}\ ,m, \end{aligned} \qquad (۴)$$

که تفاوت آن با مدل مشهور CCR (۲) در این است که در اینجا، کارآیی DMU_j نسبت به دیگران در محدوده‌ی یک یا بالاتر کمینه‌سازی می‌شود، در حالی که در آن مدل، کارآیی آن در محدوده‌ی صفر و یک بیشینه‌سازی می‌شود.

با استفاده از تبدیل Charnes و Cooper (۱۹۶۲)، برنامه‌ریزی کسری (۴) را می‌توان به طور هم‌ارز به مدل LP زیر

تبدیل کرد:

$$\begin{aligned} \min \quad & \varphi_o = \sum_{r=1}^{s} u_r y_{ro} \\ \text{s.t.} \quad & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \geq 0, \quad j = 1,\text{K}\ ,n, \\ & \sum_{i=1}^{m} v_i x_{io} = 1, \\ & u_r, v_i \geq \varepsilon, \quad r = 1,\text{K}\ ,s; \quad i = 1,\text{K}\ ,m. \end{aligned} \tag{۵}$$

زمانی که مجموعه‌ای از وزن‌های مثبت u_r^* $(r=1,\ldots,s)$ و v_i^* $(i=1,\ldots,m)$ وجود داشته باشد تا $\varphi_o^* = 1$ را تأمین کند، می‌گوییم که DMU_o ناکارآی DEA یا ناکارآی بدبینانه است. در غیر این صورت، می‌گوییم که DMU_o غیرناکارآی بدبینانه است. مدل LP (۵) نیز n بار، هر بار برای یک DMU، حل می‌شود. همه‌ی DMUهای ناکارآی بدبینانه یک مرز ناکارآ را تشکیل می‌دهند، و می‌توان در صورت لزوم آنها را برای تحلیل بیشتر با هم گروه‌بندی کرد.

اندازه‌ی عملکرد کلی: در ارزیابی قواعد انجمنی داده‌کاوی، متغیرهایی مانند *پشتیبانی*[11]، *اعتماد*[12]، *ارزش مجموعه‌ی فقرات*[13]، و *سود فروش متقابل*[14] باید ماکزیموم‌سازی شوند و می‌توان آنها را به عنوان خروجی در نظر گرفت (M.-C. Chen, 2007). برای رتبه‌بندی قواعد انجمنی کارآی خوشبینانه، Chen (۲۰۰۷) از مدل‌های رأی‌گیری استفاده کرده است، زیرا این مدل‌ها فقط داده‌های خروجی واحدها را در نظر می‌گیرند.

کارآیی‌های خوشبینانه و بدبینانه از دیدگاه‌های مختلفی اندازه‌گیری می‌شوند، که منجر به دو رتبه‌بندی متفاوت برای DMUها می‌شود. لذا یک اندازه‌ی عملکرد کلی مورد نیاز است تا رتبه‌بندی کلی DMUها به دست آید. در اینجا، ما یک اندازه‌ی عملکرد کلی جدید را برای رتبه‌بندی DMUها پیشنهاد می‌کنیم که به صورت زیر تعریف می‌شود:

$$q_j = \sqrt{\theta_j^{*2} + \varphi_j^{*2}}, \quad j = 1,\ldots,n \tag{۶}$$

که در اینجا θ_j^* و φ_j^* به ترتیب کارآیی‌های خوشبینانه و بدبینانه‌ی DMU_j هستند. روشن است که اندازه‌ی عملکرد کلی تعریف شده در (۶) بزرگی دو کارآیی را در نظر می‌گیرد.

برای راحتی، رویکردی را که عملکرد کلی هر DMU را نسبت به هر دو کارآیی خوشبینانه و بدبینانه تعیین می‌کند، رویکرد DEA با مرز دوگانه می‌نامیم. مرز تولید کارآ مجموعه‌ای از DMUهای کارآی خوشبینانه را مشخص می‌کند که عملکرد نسبتاً خوبی دارند، در حالی که مرز تولید ناکارآ مجموعه‌ای از DMUهای ناکارآی بدبینانه را مشخص می‌کند که به نسبت، عملکرد ضعیف‌تری دارند. بهترین DMU را معمولاً می‌توان از میان DMUهای کارآی خوشبینانه انتخاب کرد.

روش پیشنهادی Chen (۲۰۰۷): در اینجا، روش پیشنهادی Chen (۲۰۰۷) برای رتبه‌بندی قواعد انجمنی مورد بحث قرار می‌گیرد. در واقع، روش پیشنهادی او از یک مدل DEA که به وسیله‌ی Cook و Kress (۱۹۹۰) برای شناسایی قواعد انجمنی کارآی خوشبینانه معرفی شده است، استفاده می‌کند. این مدل به صورت زیر است:

$$\begin{aligned} \max \quad & \sum_{j=1}^{k} w_j v_{oj} \\ \text{s.t.} \quad & \sum_{j=1}^{k} w_j v_{ij} \leq 1, \quad i = 1,\ldots,m, \\ & w_j - w_{j+1} \geq d(j,\varepsilon), \quad j = 1,\ldots,k-1, \\ & w_k \geq d(k,\varepsilon). \end{aligned} \tag{۷}$$

که در اینجا w_j نشان دهنده‌ی وزن محل j-اُم است؛ v_{ij} نشان دهنده‌ی تعداد رأی‌های محل j-اُم در نامزد i (

[11] Support
[12] Confidence
[13] Itemset value
[14] Cross-selling profit

$(i=1,\ldots,m, j=1,\ldots,k)$ است، و $d(\bullet,\varepsilon)$ که تابع تشدید افتراق نامیده می‌شود، در ε نامنفی و غیرکاهشی است، و در رابطه‌ی $d(\bullet,\varepsilon)=0$ صدق می‌کند.

مدل (۷) باید برای هر نامزد o $(o=1,\ldots,m)$ حل شود. به خاطر این واقعیت که DEA غالباً چندین نامزد کارآی خوشبینانه ایجاد می‌کند (Obata & Ishii, 2003)، لذا روش پیشنهادی Chen (۲۰۰۷) از یک مدل DEAی دیگر که به وسیله‌ی Obata و Ishii (۲۰۰۳) ابداع شده است، برای افتراق قواعد انجمنی کارآی خوشبینانه استفاده می‌کند. باید توجه کرد که این مدل هیچگونه اطلاعاتی را درباره‌ی نامزدهای غیرکارآی خوشبینانه مورد استفاده قرار نمی‌دهد و فقط باید برای قواعد انجمنی کارآی خوشبینانه حل شود.

اما روش فوق‌الذکر Chen (۲۰۰۷) دارای خاصیت‌های زیر است:

- روش Chen (۲۰۰۷) نیاز به محاسبه‌ی v_{ij} از y_{ij} (خروجی‌های j -اُم از قاعده‌ی انجمنی i -اُم) دارد. گرچه آلگوریتم محاسبه‌ی v_{ij} از y_{ij} چندجمله‌ای است، ولی وقت‌گیر است. شناسایی قواعد انجمنی کارآی خوشبینانه را می‌توان از طریق روش ساده‌تر و مؤثرتری انجام داد.
- نتیجه‌ی روش Chen (۲۰۰۷) تا حد بسیار زیادی به تابع تشدید افتراق بستگی دارد.

برای نشان دادن کاربردپذیری روش پیشنهادی، مثالی از داده‌های سبد بازار از Chen (۲۰۰۷) بررسی می‌شود. قواعد انجمنی ابتدا با آلگوریتم Apriori کشف می‌شوند، که در آن حداقل پشتیبانی و حداقل اعتماد به ترتیب ۱٪ و ۱۰٪ منظور می‌شود. به این ترتیب، ۴۶ قاعده شناسایی می‌شوند که در جدول ۲ نشان داده شده‌اند. باید این قواعد انجمنی را رتبه‌بندی کنیم. تمام مدل‌ها روی یک کامپیوتر شخصی با استفاده از برنامه‌ی حل کننده‌ی LP به نام GAMS اجرا شدند. در این مثال مقدار بی‌نهایت کوچک غیرارشمیدسی $\varepsilon=10^{-10}$ منظور شده است.

جدول شماره(۲): داده‌های قواعد انجمنی.

شماره‌ی قاعده‌ی انجمنی (DMU)	خروجی‌ها			ورودی
	پشتیبانی (٪)	اعتماد (٪)	ارزش مجموعه‌ی فقرات	سود فروش متقابل
۱	۳/۸۷	۴۰/۰۹	۳۳۷/۰۰	۲۵/۶۶	۱
۲	۱/۴۲	۱۸/۱۷	۵۰۱/۰۰	۱۱/۶۳	۱
۳	۲/۸۳	۱۷/۶۴	۳۴۵/۰۰	۱۱/۲۹	۱
۴	۲/۳۴	۳۰/۸۳	۱۶۳/۰۰	۱۹/۷۳	۱
۵	۲/۶۳	۲۳/۹۰	۳۲۵/۰۰	۱۵/۳۰	۱
۶	۱/۱۹	۵۵/۶۵	۴۳۶/۰۰	۳۵/۶۱	۱
۷	۱/۱۹	۴۷/۴۲	۵۹۸/۰۰	۳۰/۳۵	۱
۸	۱/۱۹	۱۵/۷۰	۴۳۶/۰۰	۵۲/۹۱	۱
۹	۱/۱۹	۱۰/۸۲	۵۹۸/۰۰	۳۶/۴۵	۱
۱۰	۱/۱۹	۱۲/۳۲	۴۳۶/۰۰	۲۰/۰۸	۱
۱۱	۱/۱۹	۱۲/۳۲	۵۹۸/۰۰	۴۰/۰۴	۱
۱۲	۳/۸۷	۳۸/۰۸	۳۳۷/۰۰	۱۰۳/۹۷	۱
۱۳	۱/۱۸	۱۵/۰۹	۷۱۰/۰۰	۴۱/۱۹	۱
۱۴	۲/۴۴	۱۵/۲۲	۵۵۴/۰۰	۴۱/۵۶	۱
۱۵	۲/۱۴	۲۸/۲۱	۳۷۲/۰۰	۷۷/۰۲	۱
۱۶	۲/۵۱	۲۲/۸۱	۵۳۴/۰۰	۶۲/۲۶	۱
۱۷	۱/۱۹	۵۰/۹۲	۴۳۶/۰۰	۱۳۹/۰۲	۱
۱۸	۱/۱۹	۴۵/۲۵	۵۹۸/۰۰	۱۲۳/۵۲	۱
۱۹	۱/۱۹	۱۱/۷۰	۴۳۶/۰۰	۴۳/۵۴	۱
۲۰	۱/۱۹	۱۱/۷۰	۵۹۸/۰۰	۶۲/۵۰	۱

۲۱	۱/۴۲	۱۳/۹۹	۵۰۱/۰۰	۶۱/۱۶	۱
۲۲	۱/۱۸	۱۲/۲۳	۷۱۰/۰۰	۵۳/۴۵	۱
۲۳	۱/۵۰	۱۳/۶۴	۶۹۸/۰۰	۵۹/۵۹	۱
۲۴	۲/۸۳	۲۷/۸۲	۳۴۵/۰۰	۷۸/۱۷	۱
۲۵	۲/۴۴	۲۵/۲۷	۵۵۴/۰۰	۷۱/۰۰	۱
۲۶	۱/۲۵	۱۵/۹۷	۷۱۸/۰۰	۴۴/۸۷	۱
۲۷	۱/۲۲	۳۴/۸۹	۳۳۹/۰۰	۹۸/۰۴	۱
۲۸	۱/۳۰	۳۵/۱۲	۴۳۵/۰۰	۹۸/۶۸	۱
۲۹	۱/۴۲	۳۳/۸۱	۵۳۴/۰۰	۹۵/۰۱	۱
۳۰	۱/۹۱	۲۵/۲۶	۳۸۰/۰۰	۷۰/۹۷	۱
۳۱	۱/۴۳	۳۷/۱۴	۶۱۸/۰۰	۱۰۴/۳۵	۱
۳۲	۲/۳۸	۲۱/۶۳	۵۴۲/۰۰	۶۰/۷۸	۱
۳۳	۱/۱۸	۳۰/۲۴	۳۶۶/۰۰	۸۴/۹۸	۱
۳۴	۱/۲۳	۲۹/۳۶	۶۲۶/۰۰	۸۲/۵۱	۱
۳۵	۱/۵۸	۲۲/۶۵	۳۵۴/۰۰	۶۳/۶۴	۱
۳۶	۲/۳۴	۲۲/۹۹	۱۶۳/۰۰	۲۲/۷۶	۱
۳۷	۲/۱۴	۲۲/۱۴	۳۷۲/۰۰	۲۱/۹۲	۱
۳۸	۱/۹۱	۱۱/۹۴	۳۸۰/۰۰	۱۱/۸۲	۱
۳۹	۲/۰۳	۱۸/۴۲	۳۶۰/۰۰	۱۸/۲۳	۱
۴۰	۱/۱۹	۳۰/۷۳	۴۳۶/۰۰	۳۰/۴۳	۱
۴۱	۲/۶۳	۲۵/۸۷	۳۲۵/۰۰	۶۷/۵۲	۱
۴۲	۲/۵۱	۲۵/۹۸	۵۳۴/۰۰	۶۷/۸۱	۱
۴۳	۱/۵۰	۱۹/۱۶	۶۹۸/۰۰	۵۰/۰۲	۱
۴۴	۲/۳۸	۱۴/۸۵	۵۴۲/۰۰	۳۸/۷۵	۱
۴۵	۲/۰۳	۲۶/۷۳	۳۶۰/۰۰	۶۹/۷۸	۱
۴۶	۱/۱۹	۳۰/۷۳	۵۹۸/۰۰	۸۰/۲۲	۱

با اجرای مدل‌های DEAی (۳) و (۵) به ترتیب برای هر قاعده‌ی انجمنی، کارآیی‌های خوشبینانه و بدبینانه‌ی ۴۶ قاعده‌ی انجمنی را به دست می‌آوریم، و سپس آنها را با استفاده از اندازه‌ی (۶) تجمعی می‌کنیم تا نمره‌ی عملکرد کلی هر قاعده‌ی انجمنی به دست آید. نتایج در جدول ۳ نشان داده شده است، که با مراجعه به آن، مشاهده می‌شود که قاعده‌ی انجمنی شماره‌ی ۱۲ بهترین عملکرد کلی را دارد. این نشان می‌دهد که اندازه‌ی عملکرد کلی تعریف شده با معادله‌ی (۶) توجیه خوبی برای رتبه‌بندی قاعده‌ی انجمنی بدون نیاز به محاسبات زیاد است. این بزرگ‌ترین مزیت DEA با مرز دوگانه نسبت به روش‌های دیگر برای رتبه‌بندی قاعده‌ی انجمنی است. مزیت دیگر روش پیشنهادی ما آن است که این روش می‌تواند همه‌ی قواعد انجمنی را رتبه‌بندی کند.

جدول شماره(۳): ارزیابی ۴۶ قاعده‌ی انجمنی با استفاده از DEA با مرز دوگانه.

شماره‌ی قاعده‌ی انجمنی (DMU)	کارآیی خوشبینانه	کارآیی بدبینانه	عملکرد کلی	رتبه‌بندی
۱	۱/۰۰۰۰	۱/۵۱۲۹	۱/۸۱۳۶	۵
۲	۰/۷۷۹۴	۱/۰۰۰۰	۱/۲۶۷۹	۳۶
۳	۰/۸۳۷۷	۱/۰۰۰۰	۱/۳۰۴۵	۳۵
۴	۰/۷۱۰۰	۱/۰۰۰۰	۱/۲۲۶۴	۴۱
۵	۰/۷۸۳۲	۱/۰۸۸۵	۱/۳۴۱۰	۳۰
۶	۱/۰۰۰۰	۱/۰۰۲۳	۱/۴۱۵۸	۲۱

۷	۱/۰۰۰۰	۱/۰۰۴۱	۱/۴۱۷۱	۲۰
۸	۰/۶۸۵۴	۱/۰۰۲۲	۱/۲۱۴۲	۴۳
۹	۰/۸۴۷۴	۱/۰۰۰۰	۱/۳۱۰۸	۳۴
۱۰	۰/۶۶۶۲	۱/۰۰۰۰	۱/۲۰۱۶	۴۵
۱۱	۰/۸۴۷۴	۱/۰۰۴۹	۱/۳۱۴۵	۳۲
۱۲	۱/۰۰۰۰	۱/۷۴۸۰	۲/۰۱۳۸	۱
۱۳	۰/۹۸۸۹	۱/۰۰۰۰	۱/۴۰۶۴	۲۶
۱۴	۰/۹۹۹۱	۱/۳۰۰۵	۱/۶۳۹۹	۸
۱۵	۰/۷۸۲۳	۱/۲۶۴۰	۱/۴۸۶۵	۱۵
۱۶	۰/۹۸۹۰	۱/۵۲۹۰	۱/۸۲۱۰	۴
۱۷	۱/۰۰۰۰	۱/۰۰۸۵	۱/۴۲۰۲	۱۸
۱۸	۱/۰۰۰۰	۱/۰۰۸۵	۱/۴۲۰۲	۱۸
۱۹	۰/۶۶۷۷	۱/۰۰۰۰	۱/۲۰۲۴	۴۴
۲۰	۰/۸۷۸۷	۱/۰۰۳۵	۱/۳۳۳۸	۳۱
۲۱	۰/۷۹۴۵	۱/۱۶۶۶	۱/۴۱۱۵	۲۴
۲۲	۱/۰۰۰۰	۱/۰۰۰۰	۱/۴۱۴۲	۲۳
۲۳	۱/۰۰۰۰	۱/۲۳۹۲	۱/۵۹۲۴	۱۱
۲۴	۰/۸۳۷۷	۱/۴۱۵۳	۱/۶۴۴۶	۷
۲۵	۰/۹۹۹۱	۱/۵۸۴۳	۱/۸۷۳۰	۲
۲۶	۱/۰۰۰۰	۱/۰۵۸۷	۱/۴۵۶۳	۱۶
۲۷	۰/۷۵۱۸	۱/۰۰۰۰	۱/۲۵۱۱	۳۷
۲۸	۰/۸۰۷۴	۱/۱۰۱۷	۱/۳۶۵۹	۲۹
۲۹	۰/۸۹۶۱	۱/۲۰۳۴	۱/۵۰۰۴	۱۴
۳۰	۰/۷۵۰۴	۱/۱۹۵۴	۱/۴۱۱۴	۲۵
۳۱	۱/۰۰۰۰	۱/۲۱۱۹	۱/۵۷۱۲	۱۲
۳۲	۰/۹۷۶۳	۱/۴۹۹۵	۱/۷۸۹۳	۶
۳۳	۰/۶۹۶۲	۱/۰۰۰۰	۱/۲۱۸۵	۴۲
۳۴	۰/۹۵۷۸	۱/۰۴۲۴	۱/۴۱۵۶	۲۲
۳۵	۰/۶۷۰۸	۱/۰۵۳۶	۱/۲۴۹۱	۳۹
۳۶	۰/۶۰۴۷	۱/۰۰۰۰	۱/۱۶۸۶	۴۶
۳۷	۰/۷۵۳۴	۱/۱۸۳۷	۱/۴۰۳۱	۲۷
۳۸	۰/۷۲۴۲	۱/۰۰۰۰	۱/۲۳۴۷	۴۰
۳۹	۰/۷۲۲۳	۱/۰۹۴۲	۱/۳۱۱۱	۳۳
۴۰	۰/۷۴۷۴	۱/۰۰۱۵	۱/۲۴۹۷	۳۸
۴۱	۰/۷۸۳۲	۱/۳۲۱۴	۱/۵۳۶۱	۱۳
۴۲	۰/۹۸۹۰	۱/۵۷۹۷	۱/۸۶۳۸	۳
۴۳	۱/۰۰۰۰	۱/۲۶۶۵	۱/۶۱۳۷	۹
۴۴	۰/۹۷۶۳	۱/۲۶۹۴	۱/۶۰۱۴	۱۰
۴۵	۰/۷۴۹۰	۱/۲۰۸۸	۱/۴۲۲۰	۱۷
۴۶	۰/۹۲۵۹	۱/۰۰۸۵	۱/۳۶۹۱	۲۸

همچنین، نتایج حاصل از روش رتبه‌بندی Chen (۲۰۰۷) در جدول ۴ گزارش شده است. با توجه به کارآیی‌های به دست آمده در جدول ۳، واضح است که نتایج روش رتبه‌بندی Chen (۲۰۰۷) درست نمی‌باشد. در رویکرد رتبه‌بندی ما، قاعده‌ی انجمنی شماره‌ی ۱۲ در رتبه‌ی اول قرار دارد. اما در روش رتبه‌بندی Chen (۲۰۰۷) قاعده‌ی انجمنی شماره‌ی ۱ در رتبه‌ی اول قراردارد.

با توجه به داده‌های خروجی‌های موجود در جدول ۲، می‌توان دید که قاعده‌ی انجمنی شماره‌ی ۱۲ از برتری مطلق در خروجی چهارم (یعنی سود فروش متقابل) نسبت به قاعده‌ی انجمنی شماره‌ی ۱ برخوردار است (سایر خروجی‌ها یکسان می‌باشند). بنابراین، می‌بایستی در رتبه‌ی ۱ قرار بگیرد. در واقع ۶ قاعده‌ی انجمنی یعنی قواعد انجمنی شماره‌های ۶، ۷، ۱۷، ۱۸، ۲۲ و ۲۶ بر اساس رتبه‌بندی DEA با مرز دوگانه جزء ۱۱ قاعده‌ی انجمنی برتر نمی‌باشند. این نکته نیز شایان توجه است که همیشه نمی‌توان همه‌ی واحدهای کارآی خوشبینانه را بالاتر از واحدهای غیرکارآی خوشبینانه رتبه‌بندی کرد. این امر بر اساس تحقیقات قبلی صحیح است (F.-H. F. Liu & Hsuan Peng, 2008; Sinuany-Friedman & Sinuany-Stern, 1997 ; Stern & Friedman, 1998; Y.-M. Wang, Luo, & Liang, 2009).

قاعده‌ی انجمنی شماره‌ی ۲۲ هم روی مرز تولید کارآ و هم روی مرز تولید ناکارآ می‌باشد؛ یعنی هم کارآی خوشبینانه و هم ناکارآی بدبینانه است. به عبارت دیگر، دو مرز همزمان از این DMU عبور می‌کند. می‌توان آن را به صورت زیر تفسیر کرد: گرچه واحدهای کارآی خوشبینانه عملکرد خوبی دارند، ولی برخی واحدهای کارآی خوشبینانه عملکردی بدتر از سایرین دارند. به همین ترتیب، با آنکه انتظار داریم که واحدهای ناکارآی بدبینانه عملکرد ضعیفی داشته باشند، ولی برخی واحدهای ناکارآی بدبینانه بهتر از بقیه عمل می‌کنند. لذا اگر یک DMU هم کارآی خوشبینانه و هم ناکارآی بدبینانه باشد، معنایش این است که عملکرد آن نه بهترین است، و نه بدترین، مانند قاعده‌ی انجمنی شماره‌ی ۲۲ در این مثال.

جدول شماره(۴): رتبه‌بندی ۱۱ قاعده‌ی انجمنی کارآی خوشبینانه با استفاده از روش رتبه‌بندی Chen (۲۰۰۷).

شماره‌ی قاعده‌ی انجمنی (DMU)	رتبه‌بندی
۱	۱
۱۲	۲
۳۱	۳
۴۳	۴
۶	۵
۲۳	۶
۷	۷
۱۷	۸
۱۸	۹
۲۲	۱۰
۲۶	۱۱

۳- نتایج و بحث

مقبولیت داده‌کاوی با سرعت سرسام‌آوری رو به افزایش است. با استفاده از این تکنیک‌ها، قواعد مختلفی را می‌توان به دست آورد، در حالی که تنها تعداد معدودی از این قواعد، به علت محدودیت بودجه و منابع، امکان پیاده‌سازی خواهند داشت. در مقاله‌ی حاضر، ما یک رویکرد DEA با مرز دوگانه را برای رتبه‌بندی قواعد انجمنی معرفی کردیم. رویکرد DEA با مرز دوگانه نه فقط کارآیی‌های خوشبینانه‌ی قواعد انجمنی را در نظر می‌گیرد، بلکه کارآیی‌های بدبینانه‌ی آنها را نیز همزمان در نظر می‌گیرد. همچنین، ما یک اندازه‌ی جدید عملکرد کلی را برای تجمیع دو کارآیی مختلف هر قاعده‌ی انجمنی پیشنهاد کردیم. اندازه‌ی عملکرد کلی، بزرگی دو کارآیی مختلف را در نظر می‌گیرد. رویکرد DEA با مرز دوگانه با یک مثال عددی برای نشان دادن سادگی و سودمندی آن در رتبه‌بندی قواعد انجمنی بررسی شد. این روش می‌تواند کارآترین قاعده‌ی انجمنی را به آسانی و به درستی شناسایی کند. انتظار می‌رود که رویکرد DEAی پیشنهادی می‌تواند نقش مهمی در رتبه‌بندی قواعد انجمنی داشته باشد، و کاربردهای بیشتری در آینده داشته باشد. کاربردهای آن محدود به رتبه‌بندی قواعد انجمنی نیست، بلکه شامل عرصه‌های دیگر نیز هست که در آن باید شاخص‌های متعدد در فرآیند تصمیم‌گیری در نظر گرفته شوند.

٤- منابع

1. Agrawal, Rakesh, Imieli, Tomasz, & Swami, Arun. (1993). *Mining association rules between sets of items in large databases*. Paper presented at the Proceedings of the 1993 ACM SIGMOD international conference on Management of data, Washington, D.C., USA.
2. Agrawal, Rakesh, & Srikant, Ramakrishnan. (1994). *Fast Algorithms for Mining Association Rules in Large Databases*. Paper presented at the Proceedings of the 20th International Conference on Very Large Data Bases.
3. Amirteimoori, Alireza. (2007). DEA efficiency analysis: Efficient and anti-efficient frontier. *Applied Mathematics and Computation, 186*(1), 10-16. doi: http://dx.doi.org/10.1016/j.amc.2006.07.006
4. Amirteimoori, Alireza, Emrouznejad, Ali, & Khoshandam, Leila. (2013). Classifying flexible measures in data envelopment analysis: A slack-based measure. *Measurement, 46*(10), 4100-4107. doi: http://dx.doi.org/10.1016/j.measurement.2013.08.019
5. Amirteimoori, Alireza, Kordrostami, Sohrab, & Azizi, Hossein. (2016). Additive models for network data envelopment analysis in the presence of shared resources. *Transportation Research Part D: Transport and Environment, 48*, 411-424. doi: 10.1016/j.trd.2015.12.016
6. Archak, Nikolay, Ghose, Anindya, & Ipeirotis, Panagiotis G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science, 57*(8), 1485-1509. doi: 10.1287/mnsc.1110.1370
7. Azizi, Hossein. (2011). The interval efficiency based on the optimistic and pessimistic points of view. *Applied Mathematical Modelling, 35*(5), 2384-2393. doi: http://dx.doi.org/10.1016/j.apm.2010.11.055
8. Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science, 30*(9), 1078-1092. doi: doi:10.1287/mnsc.30.9.1078
9. Bellazzi, Riccardo, & Zupan, Blaz. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics, 77*(2), 81-97. doi: http://dx.doi.org/10.1016/j.ijmedinf.2006.11.006
10. Breese, John S., Heckerman, David, & Kadie, Carl. (1998). *Empirical analysis of predictive algorithms for collaborative filtering*. Paper presented at the Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, Madison, Wisconsin.
11. Camanho, A. S., & Dyson, R. G. (2005). Cost efficiency measurement with price uncertainty: a DEA application to bank branch assessments. *European Journal of Operational Research, 161*(2), 432-446. doi: http://dx.doi.org/10.1016/j.ejor.2003.07.018
12. Cao, Xin, Cong, Gao, & Jensen, Christian S. (2010). Mining significant semantic locations from GPS data. *Proceedings of the VLDB Endowment, 3*(1-2), 1009-1020. doi: 10.14778/1920841.1920968
13. Chadwick, Andrew, & May, Christopher. (2003). Interaction between States and Citizens in the Age of the Internet: "e-Government" in the United States, Britain, and the European Union. *Governance, 16*(2), 271-300. doi: 10.1111/1468-0491.00216
14. Charnes, A., & Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly, 9*(3-4), 181-186. doi: 10.1002/nav.3800090303
15. Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429-444. doi: http://dx.doi.org/10.1016/0377-2217(78)90138-8
16. Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: a

general framework and some examples. *Computer, 37*(4), 50-56. doi: 10.1109/mc.2004.1297301
17. Chen, Hsinchun, Chung, Wingyan, Qin, Yi, Chau, Michael, Xu, Jennifer Jie, Wang, Gang, . . . Atabakhsh, Homa. (2003). *Crime data mining: an overview and case studies*. Paper presented at the Proceedings of the 2003 annual national conference on Digital government research, Boston, MA, USA.
18. Chen, Min. (2013). Towards smart city: M2M communications with software agent intelligence. *Multimedia Tools and Applications, 67*(1), 167-178. doi: 10.1007/s11042-012-1013-4
19. Chen, Min. (2014). NDNC-BAN: Supporting rich media healthcare services via named data networking in cloud-assisted wireless body area networks. *Information Sciences, 284*, 142-156. doi: http://dx.doi.org/10.1016/j.ins.2014.06.023
20. Chen, Min, Gonzalez, Sergio, Leung, Victor, Zhang, Qian, & Li, Ming. (2010). A 2G-RFID-based e-healthcare system. *IEEE Wireless Communications, 17*(1), 37-43. doi: 10.1109/mwc.2010.5416348
21. Chen, Min, Ma, Yujun, Jialun, Wang, Dung Ong, Mau, & Song, Enmin. (2013). *Enabling comfortable sports therapy for patient: A novel lightweight durable and portable ECG monitoring system*.
22. Chen, Min, Mau, Dung Ong, Wang, Xiaofei, & Wang, Honggang. (2013). *The virtue of sharing: Efficient content delivery in Wireless Body Area Networks for ubiquitous healthcare*.
23. Chen, Mu-Chen. (2007). Ranking discovered rules from data mining with multiple criteria by data envelopment analysis. *Expert Systems with Applications, 33*(4), 1110-1116. doi: http://dx.doi.org/10.1016/j.eswa.2006.08.007
24. Chen, Xiaogang, Skully, Michael, & Brown, Kym. (2005). Banking efficiency in China: Application of DEA to pre- and post-deregulation eras: 1993–2000. *China Economic Review, 16*(3), 229-245. doi: http://dx.doi.org/10.1016/j.chieco.2005.02.001
25. Choi, Duke Hyun, Ahn, Byeong Seok, & Kim, Soung Hie. (2005). Prioritization of association rules in data mining: Multiple criteria decision approach. *Expert Systems with Applications, 29*(4), 867-878. doi: http://dx.doi.org/10.1016/j.eswa.2005.06.006
26. Cook, Wade D., & Kress, Moshe. (1990). A Data Envelopment Model for Aggregating Preference Rankings. *Management Science, 36*(11), 1302-1310. doi: 10.1287/mnsc.36.11.1302
27. Du, Xiao Fang, Leung, Stephen C. H., Zhang, Jin Long, & Lai, K. K. (2013). Demand forecasting of perishable farm products using support vector machine. *International Journal of Systems Science, 44*(3), 556-567. doi: 10.1080/00207721.2011.617888
28. Duan, L., Street, W. N., & Xu, E. (2011). Healthcare information systems: data mining methods in the creation of a clinical recommender system. *Enterprise Information Systems, 5*(2), 169-181. doi: 10.1080/17517575.2010.541287
29. Edirisinghe, N. C. P., & Zhang, X. (2007). Generalized DEA model of fundamental analysis and its application to portfolio optimization. *Journal of Banking & Finance, 31*(11), 3311-3335. doi: http://dx.doi.org/10.1016/j.jbankfin.2007.04.008
30. Elgendy, Nada, & Elragal, Ahmed. (2014). Big Data Analytics: A Literature Review Paper. In P. Perner (Ed.), *Advances in Data Mining. Applications and Theoretical Aspects: 14th Industrial Conference, ICDM 2014, St. Petersburg, Russia, July 16-20, 2014. Proceedings* (pp. 214-227). Cham: Springer International Publishing.
31. Ertay, Tijen, Ruan, Da, & Tuzkaya, Umut Rıfat. (2006). Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing systems. *Information Sciences, 176*(3), 237-262. doi: http://dx.doi.org/10.1016/j.ins.2004.12.001
32. Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General), 120*(3), 253-290. doi: 10.2307/2343100

33. Friedman, Lea, & Sinuany-Stern, Zilla. (1997). Scaling units via the canonical correlation analysis in the DEA context. *European Journal of Operational Research, 100*(3), 629-637. doi: 10.1016/s0377-2217(97)84108-2
34. Gavalas, Damianos, Konstantopoulos, Charalampos, Mastakas, Konstantinos, & Pantziou, Grammati. (2014). Mobile recommender systems in tourism. *Journal of Network and Computer Applications, 39*, 319-333. doi: http://dx.doi.org/10.1016/j.jnca.2013.04.006
35. Guy, Ido. (2014). *Tutorial on social recommender systems*. Paper presented at the Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea.
36. Heer, Jeffrey, & Chi, Hsin-Chou. (2001). *Identification of Web User Traffic Composition using Multi-Modal Clustering and Information Scent.* Paper presented at the Conference on Data Mining.
37. Helbig, Natalie, Ramón Gil-García, J., & Ferro, Enrico. (2009). Understanding the complexity of electronic government: Implications from the digital divide literature. *Government Information Quarterly, 26*(1), 89-97. doi: http://dx.doi.org/10.1016/j.giq.2008.05.004
38. Hsieh, Nan-Chen, & Hung, Lun-Ping. (2010). A data driven ensemble classifier for credit scoring analysis. *Expert Systems with Applications, 37*(1), 534-545. doi: http://dx.doi.org/10.1016/j.eswa.2009.05.059
39. Huang, Shu-Meng. (2013). A Study of the Application of Data Mining on the Spatial Landscape Allocation of Crime Hot Spots. In F. Bian, Y. Xie, X. Cui & Y. Zeng (Eds.), *Geo-Informatics in Resource Management and Sustainable Ecosystem: International Symposium, GRMSE 2013, Wuhan, China, November 8-10, 2013, Proceedings, Part I* (pp. 274-286). Berlin, Heidelberg: Springer Berlin Heidelberg.
40. Johnes, Jill. (2006). Measuring teaching efficiency in higher education: An application of data envelopment analysis to economics graduates from UK Universities 1993. *European Journal of Operational Research, 174*(1), 443-456. doi: http://dx.doi.org/10.1016/j.ejor.2005.02.044
41. Kambal, Eiman, Osman, Izzeldin, Taha, Methag, Mohammed, Noon, & Mohammed, Sara. (2013, 26-28 Aug. 2013). *Credit scoring using data mining techniques with particular reference to Sudanese banks.* Paper presented at the 2013 INTERNATIONAL CONFERENCE ON COMPUTING, ELECTRICAL AND ELECTRONIC ENGINEERING (ICCEEE).
42. Kincade, K. (1998). Data mining: digging for healthcare gold. *Insurance & Technology, 23*(2), 2-7.
43. Koh, Hian Chye, & Tan, Gerald. (2011). Data mining applications in healthcare. *Journal of Healthcare Information Management, 19*(2), 65.
44. Koh, Hian Chye, Tan, Wei Chin, & Goh, Chwee Peng. (2006). A Two-step Method to Construct Credit Scoring Models with Data Mining Techniques. *International Journal of Business and Information, 1*(1), 96–118.
45. Konstan, Joseph A., Walker, J. D., Brooks, D. Christopher, Brown, Keith, & Ekstrand, Michael D. (2014). *Teaching recommender systems at large scale: evaluation and lessons learned from a hybrid MOOC*. Paper presented at the Proceedings of the first ACM conference on Learning @ scale conference, Atlanta, Georgia, USA.
46. Lee, Hakyeon, Kim, Sang Gook, Park, Hyun-woo, & Kang, Pilsung. (2014). Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach. *Technological Forecasting and Social Change, 86*, 49-64. doi: http://dx.doi.org/10.1016/j.techfore.2013.08.020
47. Lenca, Philippe, Meyer, Patrick, Vaillant, Benoît, & Lallich, Stéphane. (2008). On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. *European Journal of Operational Research, 184*(2), 610-626. doi:

http://dx.doi.org/10.1016/j.ejor.2006.10.059
48. Liu, Bing, Hsu, Wynne, Chen, Shu, & Ma, Yiming. (2000). Analyzing the subjective interestingness of association rules. *IEEE Intelligent Systems, 15*(5), 47-55. doi: 10.1109/5254.889106
49. Liu, F.-H. F., & Hsuan Peng, H. (2008). Ranking of units on the DEA frontier with common weights. *Computers & Operations Research, 35*(5), 1624-1637. doi: 10.1016/j.cor.2006.09.006
50. Liu, Jianqi, Wan, Jiafu, He, Shenghua, & Zhang, Yanlin. (2014). E-Healthcare Supported by Big Data. *ZTE Communications, 12*(3), 46-52.
51. Liu, Jianqi, Wang, Qinruo, Wan, Jiafu, Xiong, Jianbin, & Zeng, Bi. (2013). Towards Key Issues of Disaster Aid based on Wireless Body Area Networks. *KSII Transactions on Internet and Information Systems, 7*(5), 1014-1035. doi: 10.3837/tiis.2013.05.005
52. Liu, Jun, Pan, Jianke, Wang, Yanping, Lin, Dingkun, Shen, Dan, Yang, Hongjun, . . . Cao, Xuewei. (2013). Component analysis of Chinese medicine and advances in fuming-washing therapy for knee osteoarthritis via unsupervised data mining methods. *Journal of Traditional Chinese Medicine, 33*(5), 686-691. doi: http://dx.doi.org/10.1016/S0254-6272(14)60043-1
53. Liu, Qiang, Wan, Jiafu, & Zhou, Keliang. (2014). Cloud Manufacturing Service System for Industrial-Cluster-Oriented Application. *Journal of Internet Technology, 15*(3), 373-380. doi: 10.6138/JIT.2014.15.3.06
54. Liu, Shiang-Tai. (2008). A fuzzy DEA/AR approach to the selection of flexible manufacturing systems. *Computers & Industrial Engineering, 54*(1), 66-76. doi: http://dx.doi.org/10.1016/j.cie.2007.06.035
55. Lu, Chi-Jie, & Wang, Yen-Wen. (2010). Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting. *International Journal of Production Economics, 128*(2), 603-613. doi: http://dx.doi.org/10.1016/j.ijpe.2010.07.004
56. Maaß, Dennis, Spruit, Marco, & de Waal, Peter. (2014). Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques. *Decision Analytics, 1*(1), 4. doi: 10.1186/2193-8636-1-4
57. Mannino, Michael, Hong, Sa Neung, & Choi, In Jun. (2008). Efficiency evaluation of data warehouse operations. *Decision Support Systems, 44*(4), 883-898. doi: http://dx.doi.org/10.1016/j.dss.2007.10.011
58. Ng, Raymond T., Lakshmanan, Laks V. S., Han, Jiawei, & Pang, Alex. (1998). *Exploratory mining and pruning optimizations of constrained associations rules*. Paper presented at the Proceedings of the 1998 ACM SIGMOD international conference on Management of data, Seattle, Washington, USA.
59. Obata, Tsuneshi, & Ishii, Hiroaki. (2003). A method for discriminating efficient candidates with ranked voting data. *European Journal of Operational Research, 151*(1), 233-237. doi: http://dx.doi.org/10.1016/S0377-2217(02)00597-0
60. Olafsson, Sigurdur, Li, Xiaonan, & Wu, Shuning. (2008). Operations research and data mining. *European Journal of Operational Research, 187*(3), 1429-1448. doi: http://dx.doi.org/10.1016/j.ejor.2006.09.023
61. Padhy, Neelamadhab, Mishra, Pragnyaban, & Panigrahi, Rasmita. (2012). The Survey of Data Mining Applications and Feature Scope. *International Journal of Computer Science, Engineering and Information Technology, 2*(3), 43-58. doi: 10.5121/ijcseit.2012.2303
62. Peng, Yi, Zhang, Yong, Tang, Yu, & Li, Shiming. (2011). An incident information management framework based on data integration, data mining, and multi-criteria decision making. *Decision Support Systems, 51*(2), 316-327. doi: http://dx.doi.org/10.1016/j.dss.2010.11.025

63. Resnick, Paul, & Varian, Hal R. (1997). Recommender systems. *Communications of the ACM, 40*(3), 56-58. doi: 10.1145/245108.245121
64. Schroedl, Stefan, Wagstaff, Kiri, Rogers, Seth, Langley, Pat, & Wilson, Christopher. (2004). Mining GPS Traces for Map Refinement. *Data Mining and Knowledge Discovery, 9*(1), 59-87. doi: 10.1023/b:dami.0000026904.74892.89
65. Shafer, Scott M., & Byrd, Terry A. (2000). A framework for measuring the efficiency of organizational investments in information technology using data envelopment analysis. *Omega, 28*(2), 125-141. doi: http://dx.doi.org/10.1016/S0305-0483(99)00039-0
66. Shyam, Varan Nath. (2006). *Crime Pattern Detection Using Data Mining*. Paper presented at the Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology.
67. Silver, M., Sakata, T., Su, H. C, Herman, C, Dolins, S. B., & O'Shea, M. J. (2001). Case study: how to apply data mining techniques in a healthcare data warehouse. *Journal of Healthcare Information Management, 15*(2), 155-164.
68. Sinuany-Stern, Zilla, & Friedman, Lea. (1998). DEA and the discriminant analysis of ratios for ranking units. *European Journal of Operational Research, 111*(3), 470-478. doi: 10.1016/s0377-2217(97)00313-5
69. Srikant, Ramakrishnan, Vu, Quoc, & Agrawal, Rakesh. (1997). *Mining association rules with item constraints*. Paper presented at the Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, Newport Beach, CA.
70. Sullivan, Brooke, & Mitra, Sinjini. (2014). Community Issues in American Metropolitan Cities. *Journal of Cases on Information Technology, 16*(1), 23-39. doi: 10.4018/jcit.2014010103
71. Sun, Jimeng, & Reddy, Chandan K. (2013). *Big Data Analytics for Healthcare.* Paper presented at the in Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Austin.
72. Tan, Pang-Ning, Kumar, Vipin, & Srivastava, Jaideep. (2002). *Selecting the right interestingness measure for association patterns*. Paper presented at the Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, Edmonton, Alberta, Canada.
73. Tejeda-Lorente, A., Bernabé-Moreno, J., Porcel, C., & Herrera-Viedma, E. (2014). Integrating Quality Criteria in a Fuzzy Linguistic Recommender System for Digital Libraries. *Procedia Computer Science, 31*, 1036-1043. doi: http://dx.doi.org/10.1016/j.procs.2014.05.357
74. Thornton, Dallas, Mueller, Roland M., Schoutsen, Paulus, & van Hillegersberg, Jos. (2013). Predicting Healthcare Fraud in Medicaid: A Multidimensional Data Model and Analysis Techniques for Fraud Detection. *Procedia Technology, 9*, 1252-1264. doi: http://dx.doi.org/10.1016/j.protcy.2013.12.140
75. Wan, Jiafu, Li, Di, Zou, Caifeng, & Zhou, Keliang. (2012). M2M Communications for Smart City: An Event-Based Architecture. 895-900. doi: 10.1109/cit.2012.188
76. Wan, Jiafu, Zhang, Daqiang, Zhao, Shengjie, Yang, Laurence, & Lloret, Jaime. (2014). Context-aware vehicular cyber-physical systems with cloud support: architecture, challenges, and solutions. *IEEE Communications Magazine, 52*(8), 106-113. doi: 10.1109/mcom.2014.6871677
77. Wan, Jiafu, Zou, Caifeng, Ullah, Sana, Lai, Chin-Feng, Zhou, Ming, & Wang, Xiaofei. (2013). Cloud-enabled wireless body area networks for pervasive healthcare. *IEEE Network, 27*(5), 56-61. doi: 10.1109/mnet.2013.6616116
78. Wang, Gang, Chen, Hsinchun, & Atabakhsh, Homa. (2004). Automatically detecting deceptive criminal identities. *Communications of the ACM, 47*(3), 70-76. doi: 10.1145/971617.971618
79. Wang, Ying-Ming, Luo, Ying, & Liang, Liang. (2009). Ranking decision making units by imposing a minimum weight restriction in the data envelopment analysis. *Journal of*

Computational and Applied Mathematics, *223*(1), 469-484. doi: http://dx.doi.org/10.1016/j.cam.2008.01.022

80. Wang, Ying-Ming, & Yang, Jian-Bo. (2007). Measuring the performances of decision-making units using interval efficiencies. *Journal of Computational and Applied Mathematics*, *198*(1), 253-267. doi: http://dx.doi.org/10.1016/j.cam.2005.12.025
81. Zheng, Yu, Zhang, Lizhu, Wie, Xing, & Ma, Wei-Ying. (2009). *Mining interesting locations and travel sequences from GPS trajectories*. Paper presented at the Proceedings of the 18th international conference on World wide web, Madrid, Spain.