



واده کاوی

مہائیم و کار بردہ

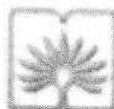
مصطفی برومندزادہ ر محمد مہدی لطفی نژاد

عضو هیات علمی دانشکدہ پیام نور



انتشارات مینوفر
Minufar.ir

مشہد مقدس - ۱۳۹۵



سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران

- سرشناسه : برومندزاده، مصطفی، ۱۳۵۷
عنوان و نام پدیدآور : داده‌کاوی، مفاهیم و کاربردها / مصطفی برومندزاده و محمدمهدی لطفی‌نژاد.
مشخصات نشر : مشهد: مینوفر، ۱۳۹۵.
مشخصات ظاهری : ۱۰۸ ص.: مصور (بخشی رنگی)، جدول، نمودار.
فروست : فناوری؛ ۸۱۰۳.
شابک : ۱۳۰۰۰۰ ریال-۹۳-۸۰۶۰-۶۰۰-۹۷۸ :
وضعیت فهرست نویسی : فیبا
یادداشت : کتابنامه: ص. [۱۰۵] - ۱۰۸.
موضوع : داده‌کاوی
موضوع : Data mining
شناسه افزوده : لطفی‌نژاد، محمدمهدی، ۱۳۶۱ -
رده بندی کنگره : ۹/۷۶QA ۱۳۹۵ ۴ب۲د/
رده بندی دیویی : ۰۰۶/۲۱۳ :
شماره کتابشناسی ملی : ۴۳۳۱۳۳۱ :



عنوان کتاب: داده کاوی، مفاهیم و کاربردها

نویسندگان: مصطفی برومندزاده - محمدمهدی لطفی

طراح جلد: مینوف

صفحه‌آرا: اکرم ملک‌نژاد

شمارگان: ۱۰۰۰ نسخه

نوبت چاپ: اول - ۱۳۹۵

چاپ و صحافی: دقت

قطع: وزیری

شابک: ۹۷۸-۶۰۰-۸۰۶۰-۹۳-۲

قیمت: ۱۲,۰۰۰ تومان

تمام حقوق چاپ و نشر این اثر محفوظ است.

انتشارات مینوف: مشهد مقدس، بلوار جلال آل احمد ۶۹ پ ۱۱۶، ص.پ: ۹۱۸۹۵-۱۷۵۵

همراه: ۰۹۳۵۲۱۶۳۷۵۵ - ۰۹۱۲۰۶۸۲۷۵۵ تلفن و دورنگار: ۰۵۱-۳۸۳۲۳۵۵۳

فہرست مطالب

۹..... پیشگفتار

۱۵..... فصل اول: مفاهیم داده کاوی

۱۷..... زیربنا ذخیره سازی و دستیابی اطلاعات

۱۸..... مختار بانک اطلاعاتی سازمان

۲۱..... مفاهیم پایه داده کاوی

۲۲..... تعریف داده کاوی

۲۲..... تاریخچه داده کاوی

۲۴..... مراحل فرایند کشف دانش در پایگاه داده ها

۲۵..... انبارش داده ها

۲۶..... انتخاب داده ها

۲۶..... تبدیل داده ها

۲۶..... کاوش در داده ها

۲۷..... تفسیر نتیجه

۲۷..... عملیات های داده کاوی

۲۸..... مدل سازی پیشگویی کننده

۲۹..... تقطیع پایگاه داده ها

۲۹..... تحلیل پیوند

۳۰..... تشخیص انحراف

۳۰..... الگوریتم های داده کاوی

۳۱..... شبکه های عصبی

۳۲..... درخت های انتخاب

۳۴..... استنتاج قانون

۳۴..... الگوریتم های ژنتیک

۳۵	مدل فرآیند دوسویه
۳۵	تعریف مسئله
۳۶	ساختن یک پایگاه داده‌ی داده‌کاوی
۳۷	جستجوی داده
۳۸	آماده‌سازی داده برای مدل‌سازی
۳۸	ساختن مدل
۳۹	تائید اعتبار ساده
۳۹	ارزیابی و تفسیر؛ تأیید اعتبار مدل
۳۹	طراحی و معماری مدل و نتایج
۴۳	سابقه داده‌کاوی
۴۵	مفهوم آلودگی عبارت داده‌کاوی
۴۵	نرم‌افزارهای داده‌کاوی
۵۶	ابزارهای تجاری داده‌کاوی: ToolsDM Commercial
۵۸	منابع اطلاعاتی مورد استفاده
۵۸	پیشرفت در تکنولوژی‌های داده‌کاوی
۵۹	دیتامارت
۶۰	انتبار داده‌ها
۶۲	عناصر داده‌کاوی
۶۴	فنون داده‌کاوی
۶۴	ابزارهای پرس‌وجو
۶۴	فنون آماری
۶۴	مصورسازی
۶۴	پردازش تحلیلی پیوسته
۶۵	یادگیری مبتنی بر مورد
۶۵	درختان تصمیم‌گیری
۶۵	قوانین وابستگی
۶۵	شبکه‌های عصبی
۶۵	الگوریتم ژنتیکی
۶۷	محدودیت‌های داده‌کاوی

۶۸.....حفاظت از حریم شخصی در سیستم‌های داده‌کاوی

۷۲.....فصل دوم: کاربردهای داده‌کاوی

۷۵.....کاربرد داده‌کاوی در کسب‌وکار هوشمند بانک

۷۷.....داده‌کاوی در مدیریت ارتباط با مشتری

۷۸.....کاربردهای داده‌کاوی در کتابخانه‌ها و محیط‌های دانشگاهی

۷۹.....کاربردهای داده‌کاوی در کتابخانه‌ها

۸۰.....داده‌کاوی و مدیریت مؤسسات دانشگاهی

۸۰.....کاربردهای داده‌کاوی در مؤسسات دانشگاهی

۸۱.....داده‌کاوی و مدیریت بهینه وب سایت‌ها

۸۲.....داده‌کاوی و مدیریت دانش

۸۳.....کاربرد داده‌کاوی در آموزش عالی

۸۶.....فصل سوم: داده‌کاوی - شهر الکترونیک

۹۱.....زمینه داده‌کاوی در شهر الکترونیک

۹۳.....کاربردهای داده‌کاوی در شهر الکترونیک

۹۴.....کشف علایق و انگیزه‌های شهروندان و تولید سرویس‌های شخصی‌سازی

۹۵.....تجدید ساختار سایت وب شهر و افزایش کارایی سیستم

۹۶.....تقویت برنامه‌ریزی‌های دولت و ترویج نوآوری

۹۷.....بهبود تحلیل‌ها و تصمیمات دولت

۹۸.....چالش‌های داده‌کاوی در شهر الکترونیک

۹۸.....کیفیت داده‌ها

۹۹.....قابلیت انتقال داده‌ها و استفاده از اطلاعات

۱۰۰.....چالش برآورد مدل‌های داده‌کاوی

۱۰۲.....دقت نتایج در متدهای داده‌کاوی

۱۰۳.....پیچیدگی و هزینه زمانی

۱۰۴.....محرمانگی داده‌ها

۱۱۰.....مراجع



در دنیای بشماره رسانی امروز، اطلاعات بعنوان یکی از فاکتورهای تولیدی مهم پدیدار شده است. در اینجا تلاش برای استخراج اطلاعات از داده‌ها توجه بسیاری از افراد دخیل در صنعت اطلاعات و حوزه‌های وابسته را به خود جلب نموده است. با گسترش سیستم‌های اطلاعاتی و حجم بالای داده‌های ذخیره شده در این سیستم‌ها، نیاز به ابزاری است تا بتوان داده‌های ذخیره‌شده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربران قرار داد. با استفاده از ابزارهای گوناگون گزارش‌گیری، می‌توان به اطلاعاتی جهت نتیجه‌گیری در مورد داده‌ها و روابط منطقی میان آن‌ها دست یافت اما وقتی که حجم داده‌ها بالا باشد، کاربران هرچند هم زبردست و باتجربه باشند نمی‌توانند الگوهای مفید را در میان حجم انبوه داده‌ها تشخیص دهند و یا اگر هم قادر به این کار باشند، هزینه عملیات بسیار بالاست. از سوی دیگر کاربران معمولاً فرضیه‌ای را مطرح می‌کنند و سپس بر اساس گزارشات مشاهده‌شده به اثبات یا رد فرضیه می‌پردازند، درحالی‌که امروزه نیاز به روش‌هایی است که اصطلاحاً به کشف دانش بپردازند یعنی با کمترین دخالت کاربر و به‌صورت

خودکار الگوها و رابطه‌های منطقی را بیان نمایند. داده‌کاوی یکی از مهم‌ترین این روش‌ها است که به‌وسیله آن الگوهای مفید در داده‌ها با حداقل دخالت کاربران شناخته می‌شوند و اطلاعاتی را در اختیار تحلیلگران قرار می‌دهند تا بر اساس آن‌ها تصمیمات مهم و حیاتی در سازمان‌ها اتخاذ شوند. در داده‌کاوی از بخشی از علم آمار به نام تحلیل اکتشافی داده‌ها استفاده می‌شود که در آن بر کشف اطلاعات نهفته و ناشناخته از درون حجم انبوه داده‌ها تأکید می‌شود. علاوه بر این داده‌کاوی با هوش مصنوعی و یادگیری ماشین نیز ارتباط تنگاتنگی دارد، بنابراین می‌توان گفت در داده‌کاوی تئوری‌های پایگاه داده‌ها، هوش مصنوعی، یادگیری ماشین و علم آمار را در هم می‌آمیزند تا زمینه کاربردی فراهم شود.

باید توجه داشت که اصلاح داده‌کاوی زمانی به‌کاربرده می‌شود که با حجم بزرگی از داده‌ها مواجه باشیم. بر به حجم داده‌ها بیشتر و روابط میان آن‌ها پیچیده‌تر باشد دسترسی به اطلاعات نهفته در میان داده‌ها مشکل‌تر می‌شود و نقش داده‌کاوی به‌عنوان یکی از روش‌های کشف دانش، روشن‌تر می‌گردد. با گسترش فناوری اطلاعات و ارتباطات در جهان و ورود سریع آن به زندگی روزمره مردم مسائل و ضرورت‌های تازه‌ای به وجود آمده است. امروزه انسان توسعه‌یافته کسی است که به اطلاعات دسترسی داشته باشد و دسترسی به اطلاعات نه یک ضرورت، که یک قدرت محسوب می‌شود. در این میان شهرها به‌عنوان مراکز قدرت انسانی و تمدن‌های بشری بیش‌ازپیش اهمیت یافته‌اند. به اعتقاد الوین تافلر، مردم کره زمین تا به امروز سه موج اساسی تحول را پشت سر گذاشته‌اند:

موج اول، موج انقلاب کشاورزی است که زمان آغاز آن بر کسی مشخص نیست.

موج دوم، انقلاب صنعتی است که به دنبال اختراع ماشین بخار در سال ۱۷۶۴ میلادی آغاز شد.

موج سوم یا انقلاب انفورماتیک است که از سال ۱۹۴۶ میلادی که بشر به ساخت کامپیوتر نائل آمده آغاز گشته است.

اگر در موج دوم سخت‌افزارها به کمک انسان‌ها می‌آمدند، در موج سوم این نرم‌افزارها هستند که به خدمت بشر می‌شتابند و تفکرات و تصورات آدمی را به شکل کد و با کمک امواج حامل انرژی مبادله می‌کنند.

در موج سوم، انسان هر روز که بیشتر یاد می‌گیرد، بیشتر می‌فهمد که با حقیقت فاصله دارد. این موج را موج خردورزی نیز لقب داده‌اند زیرا در این عرصه‌ها، انسان‌ها دیگر فرصت ندارند زیاد با هم صحبت کنند، همه‌چیز تعریف شده و برای هر تعریف، یک کد در نظر گرفته شده است. از سوی دیگر در دنیای به شدت رقابتی امروز، اطلاعات به عنوان یکی از فاکتورهای تولیدی مهم پدیدار شده است. در نتیجه تلاش برای استخراج اطلاعات از داده‌ها توجه بسیاری از افراد دخیل در صنعت اطلاعات و حوزه‌های وابسته را به خود جلب نموده است.

حجم بالای داده‌های دائماً در حال رشد است و تنوع آن‌ها به شکل داده متنی، اعداد، گرافیک، نقشه، عکس، تصاویر ماهواره‌ای و عکس‌های گرفته شده با اشعه ایکس نمایانگر پیچیدگی کار تبدیل داده‌ها به اطلاعات است. علاوه بر این، تفاوت وسیع در فرآیندهای تولید داده مثل مثل روش آنالوگ مبتنی بر کاغذ و روش دیجیتال مبتنی بر کامپیوتر، مزید بر علت شده است. استراتژی‌ها و فنون متعددی برای گردآوری، ذخیره، سازمان‌دهی و مدیریت کارآمد داده‌های

موجود و رسیدن به نتایج معنی دار بکار گرفته شده‌اند. بعلاوه، عملکرد مناسب ابر داده که داده‌ای درباره داده است در عمل عالی به نظر می‌رسد.

پیشرفت‌های حاصل در علم اطلاع‌رسانی و تکنولوژی اطلاعات، فنون و ابزارهای جدیدی برای غلبه بر رشد مستمر و تنوع بانک‌های اطلاعاتی تأمین می‌کنند. این پیشرفت‌ها هم در بعد سخت‌افزاری و هم نرم‌افزاری حاصل شده‌اند. ریزپردازنده‌های سریع، ابزارهای ذخیره داده‌های انبوه پیوسته و غیر پیوسته، ابرکنرها، چاپگرها و دیگر ابزارهای جانبی نمایانگر پیشرفت‌های حوزه سخت‌افزار هستند. پیشرفت‌های حاصل در نظام‌های مدیریت بانک اطلاعات در طی چهار دهه گذشته نمایانگر تلاش‌های بخش نرم‌افزاری است. این تلاش‌ها در بخش نرم‌افزار را می‌توان به‌عنوان یک حرکت پیش‌رونده از ایجاد یک بانک اطلاعات ساده، بانک‌های رابطه‌ای و سلسله مراتبی برای پاسخگویی به نیاز روزافزون سازمان‌ها و بازیابی اطلاعات ملاحظه نمود. بدین منظور در هر دوره، نظام‌های مدیریت بانک اطلاعاتی سازگار با نرم‌افزارهای سیستمی و سخت‌افزارهای رایج گسترش یافته‌اند. در این رابطه می‌توان از محصولاتی مانند، Dbase-IV, Unity, Sybase, Oracle و غیره نام برد.

داده کاوی یکی از پیشرفت‌های اخیر در فناوری مدیریت داده است. این تکنیک عظیم، مجموعه‌ای از فنون است که به شخص امکان می‌دهد تا ورای داده‌پردازی معمولی حرکت کند و به استخراج اطلاعاتی که در انبوه داده‌ها مخفی و یا پنهان است دست یابد. انگیزه برای گسترش داده کاوی به‌طور عمده از دنیای تجارت دهه ۱۹۹۰ پدید آمد. مثلاً داده کاوی در حوزه بازاریابی، به دلیل پیوستگی غیرقابل‌انتظاری که بین پروفایل یک مشتری و الگوی خرید او ایجاد می‌کند اهمیتی خاص دارد.

تحلیل رکوردهای حجیم نگهداری سخت‌افزارهای صنعتی، داده‌های هواشناسی و دیدن کانال‌های تلویزیونی از دیگر کاربردهای آن است. در حوزه مدیریت کتابخانه کاربرد داده‌کاوی به‌عنوان فرایند مآخذ کاوی نام‌گذاری شده است. این کتاب ابتدا به ساختار اولیه و پیش نیازهای ورود به تکنیکهای داده کاوی می‌پردازد؛ سپس عناصر داده‌ای توصیف می‌شوند و در پایان چگونگی به‌کارگیری داده‌کاوی مورد بحث قرار می‌گیرد.

مصطفی برومند زاده

محمد مهدی لطفی‌نژاد

www.ketab.ir