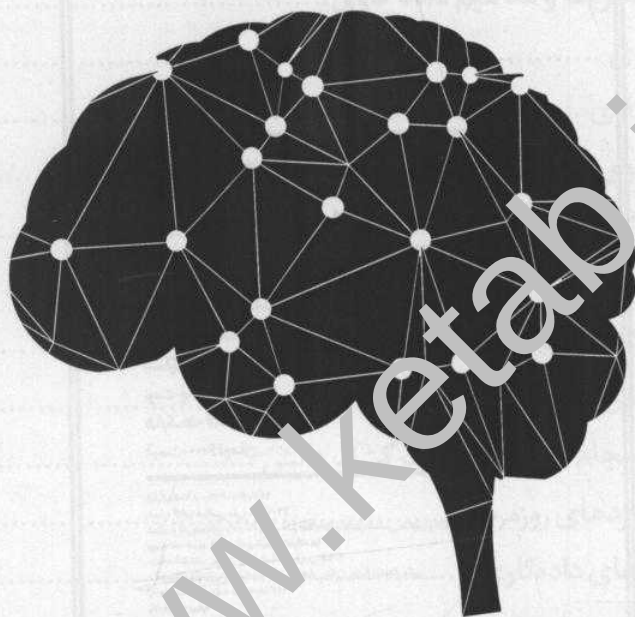




داده‌کاوی در علم اطلاعات و دانش‌شناسی مفاهیم و کاربرد

دکتر حمید قاضی‌زاده

(عضو هیأت علمی دانشگاه پیام‌نور)

مهدیه‌خزانه‌ها

داده‌کاوی در علم اطلاعات و دانش‌شناسی: مفاهیم و کاربرد

نویسنده: دکتر حمید قاضی‌زاده، مهدیه خزانه‌ها

گرافیک: عطیه سادات جعفرنیا

فردی: شادان ارشدی

شمارگان: ۰۰۰۰ نسخه

بیت ج: اول، ۱۳۹۸

شاب: ۹۷۸-۶-۲۷۶-۸۴-۵

قیمت: ۰۰ تومان

شابک: ۹۷۸-۶-۲۷۶-۸۴-۵

شماره کتاب شناسی ملی: ۹۷۸-۶-۲۷۶-۸۴-۵

عنوان و نام پدیدآور: داده کاوی در علم اطلاعات و دانش‌شناسی مفاهیم و کاربرد

نویسنده: حمید قاضی زاده، مهدیه خزانه‌ها

مشخصات نشر: تهران: چاپار/اساطیر پارس، ۱۳۹۵

مشخصات فیزیکی: ۸۲ صفحه، مصور، بخش رنگی

یادداشت: کتاب نامه ص: (۸۲-۸۱)

یادداشت: نمایه

موضوع: کتابخانه/داده پردازش

موضوع: Libraries--Data processing

موضوع: داده کاوی

موضوع: Data mining

رده بندی: ۲۵۰/۰۲۸۵

رده بندی: ۲۶۷۸/۹/۲۵۲

سرشناسه: قاضی زاده، حمید، ۱۳۵۲

وضعیت فهرست نویسی: فیا



فهرست

۸	 مقدمه
۹	 بخش اول: تعاریف و مفاهیم داده کاوی
۱۰	 تعریف داده کاوی
۱۱	 چیهستی داده کاوی
۱۳	 مفاهیم داده کاوی
۱۶	 چرا داده کاوی؟
۱۸	 پیشینه داده کاوی
۲۰	 مراحل فرایند کشف دانش زیانگه داده ها
۲۳	 الگوریتم ها و فنون داده کاوی
۲۹	 گام ها و مراحل انجام داده کاوی جهت کشف دانش
۳۵	 داده کاوی و کاربردهای روزمره آن
۴۴	 ابزارها و نرم افزارهای داده کاوی
۵۰	 وب کاوی
۵۳	 متن کاوی و کاربرد آن در سازمان دهی دانش
۶۱	 بخش دوم: داده کاوی در کتابخانه
۶۲	 داده کاوی در کتابخانه ها
۶۴	 کاربردهای داده کاوی در کتابخانه ها
۷۵	 آینده داده کاوی در کتابخانه ها
۷۸	 منابع
۸۰	 نمایه

این روزها در بیشتر سازمانها، داده‌ها و اطلاعات به سرعت در حال جمع‌آوری و ذخیره شدن می‌باشند. اما می‌توان اعا کرد که علیرغم این حجم انبوه داده‌ها، امروزه سازمانها با فقر دانش در تصمیم‌گیری روبرو هستند. در این راستا در طول دهه گذشته، با پیشرفت روزافزون کاربرد پایگاه داده‌ها، حجم داده‌های ثبت شده به طور متوسط هر ۵ سال، ۲ برابر می‌شود. پس در این میان سازمانهایی موفق اند که بتوانند حداقل اندک از داده‌های شان را تحلیل کنند. از سال ۱۹۵۰ به بعد که رایانه، در تحلیل و ذخیره‌سازی داده‌ها به سرعت در دسترس قرار گرفت، حجم اطلاعات ذخیره شده در آن پس از حدود ۲۰ سال، دو برابر شد و هم‌زمان با پیشرفت فناوری اطلاعات، حجم داده‌ها در پایگاه داده‌ها هر ساله رو به افزایش می‌باشد. در این میان ظهور شبکه جهانی وب، رشد سیستم‌های یکپارچه اطلاعاتی، سیستم‌های یکپارچه بانکی، تجارت الکترونیکی و... موجب شده تا لحظه به لحظه به حجم داده‌ها در پایگاه داده‌ها اضافه شده و انبارها (توده‌های) عظیمی از داده‌ها ایجاد شوند. ایند به طوری که ضرورت کشف و استخراج سریع و دقیق دانش از این پایگاه داده‌ها، بیش از پیش احساس می‌شود.

بنابراین بهره‌گیری از قدرت فرایند داده‌کاوی جهت شناسایی "گوشه‌ها و مدلها و نیز ارتباط عناصر مختلف در پایگاه داده جهت کشف دانش نهفته در داده‌ها و نهایتاً تبدیل داده به اطلاعات، روز به روز ضروری تر می‌شود. در داده‌کاوی معمولاً به کشف الگوهای مفید از میان داده‌ها اشاره می‌شود. منظور از الگوی مفید، مدلی در داده‌ها است که ارتباط میان یک زیرمجموعه از داده‌ها را توصیف می‌کند و معتبر، ساده، قابل فهم و جدید است. به عبارت دیگر، فن‌آوری داده‌کاوی، سازمان را قادر می‌سازد تا از سرمایه داده‌های شان بهره‌برداری نمایند. از این ابزار برای پشتیبانی فرایند تصمیم‌گیری استفاده می‌گردد. بنابراین ضروری است تا برای به کار بستن این ابزار در سازمان‌ها، اهمیت بیشتری قائل شده تا در نهایت به فرایند تصمیم‌گیری بهینه مدیران منجر شود.

در این راستا، با توجه به تحولات و تغییرات گسترده در عرصه دانش و رشد علوم در عرصه‌های مختلف، این ضرورت احساس شد که متون علمی و دوره‌های آموزشی دانشگاهها و موسسات آموزش عالی نیز، مخصوص رشته‌های دانشگاهی که به نوعی با موضوع داده و اطلاعات ارتباط تنگاتنگی دارند، دستخوش تغییر و تحول قرار گرفته و در نوع برنامه‌های علمی و واحدهای درسی خود، بازنگری مجددی به عمل آورند. از همین رو چنانچه به طور خاص نگاه خود را به برخی از رشته‌های دانشگاهی معطوف داریم، متوجه خواهیم شد که برای مثال رشته کتابداری در دانشگاه‌ها، هم از ناحیه عنوان رشته و هم از ناحیه

سرفصل، با تغییرات کلی همراه بوده است. به گونه ای که پیرو مصوبات مربوط به وزارت علوم، تحقیقات و فناوری، عنوان رشته کتابداری به رشته علم اطلاعات و دانش شناسی تغییر کرده و همچنین سرفصل آن در مقاطع کارشناسی، کارشناسی ارشد و دکتری نیز دستخوش تغییر و تحول اساسی شده است.

بنابراین با در نظر داشتن تغییرات عمده در سرفصل دروس رشته علم اطلاعات و دانش شناسی، عنوان یکی از درسها، به موضوع داده کاوی اختصاص یافته است. از همین رو، نبود منبع و کتاب مشخص در این زمینه و مخصوصاً به زبان فارسی برای دانشجویان در مقاطع تحصیلی مختلف، زمینه ای شد تا مولف برای این موضوع تأملی اندیشیده و کتاب حاضر با عنوان داده کاوی در علم اطلاعات و دانش شناسی: مفاهیم و کاربرد را به رشته تحریر درآورد. شایان ذکر است یادآور شود که مخاطبان این کتاب نه تنها دانشجویان و محققان حوزه کتابداری، اطلاع رسانی و علم اطلاعات می باشند، بلکه سایر دانشجویان و محققان رشته های علوم انسانی که قصد داشته تا چشم انداز مشخصی از داده کاوی بدست آورده و یا با مفاهیم و تعاریف اولیه مرتبط با داده کاوی آشنا شوند نیز، می توانند از جمله مخاطبان کتاب حاضر باشند.

کتاب حاضر طی دو بخش کاینده شده است. بخش اول کتاب تلاش دارد تا به تعاریف، مفاهیم، فنون و روش های داده کاوی نگاه داشته باشد و در بخش دوم کارکرد و مفاهیم داده کاوی در کتابخانه ها، مورد بررسی قرار خواهد گرفت.

شایسته است یادآوری نماید که کتاب پیش رو به منظور آشنایی بیشتر اساتید و دانشجویان رشته کتابداری، پرکردن خلاء درسی موجود در این رشته و تریب به منابعی که جهت تدوین کتب با موضوع داده کاوی بوده و با این هدف تألیف شد تا فتح بابی باشد برای چاپ کتب جدید داده کاوی به همراه محاسبات دقیق آماری، عملی و کاربردی در حوزه کتابداری، علم اطلاعات و دانش شناسی. بی شک این کتاب، خالی از نقص و ایراد نبوده و برای بهبود و لحاظ کردن اندیشه ها، نویسندگان و صاحب نظران در مراحل بعدی، نیازمند دریافت نظرها و پیشنهادهای عالمانه صاحبان این حوزه علم اطلاعات و کتابداری می باشد تا با ارائه دیدگاههای اندیشمندانه خود، مولف را در ویرایش های بعدی کتاب یاری رسانند.

دکتر حمید قاضی زاده

عضو گروه علم اطلاعات و دانش شناسی

دانشگاه پیام نور

تابستان ۱۳۹۵

در دنیای پیشرفته و الکترونیکی امروزی، با گسترش سیستم های پایگاهی و حجم بالای داده های ذخیره شده در این سیستم ها، نیاز به ابزاری است تا بتوان داده های ذخیره شده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربران قرارداد.

با استفاده از پرس و جوی داده در SQL و ابزارهای گوناگون گزارش گیری معمولی، می توان اطلاعاتی را در اختیار کاربران قرار داد تا بتوانند به نتیجه گیری در مورد داده ها و روابط منطقی میان آنها بپردازند اما وقتی که حجم داده ها بالا باشد، کاربران هر چند زیر دست و با تجربه باشند نمی توانند الگوهای مفید را در میان حجم انبوه داده ها تشخیص دهند و یا اگر قادر به این کار هم باشند، هزینه عملیات از نظر نیروی انسانی و مادی بسیار بالاست. از سوی دیگر کاربران معمولاً فرضیه ای را مطرح می کنند و سپس براساس گزارشات مشاهده شده به اثبات یا رد فرضیه می پردازند، درحالی که امروزه نیاز به روشهایی است که اصطلاحاً به کشف دانش گفته می شود، یعنی با کمترین دخالت کاربر و به صورت خودکار الگوها و رابطه های منطقی را بیان نمایند.

داده کاوی یکی از مهمترین این روشها است که به وسیله آن، الگوهای مفید در داده ها با حداقل دخالت کاربران شناخته می شوند و اطلاعاتی را در اختیار کاربران و تحلیل گران قرار می دهند تا براساس آنها تصمیمات مهم و حیاتی در سازمانها اتخاذ شوند.

طی داده کاوی از بخشی از علم آمار به نام تحلیل اکتشافی داده استفاده می شود که در آن بر کشف اطلاعات نهفته و ناشناخته از درون حجم انبوه داده ها تاکید می شود. علاوه بر این، داده کاوی، با هوش مصنوعی و یادگیری ماشین نیز ارتباط تنگاتنگی دارد، بنابراین می توان گفت، داده کاوی، تئوری پایگاه داده ها، هوش مصنوعی، یادگیری ماشین و علم آمار را در هم می آمیزند تا زمینه کاربردی فراهم شود.

باید توجه داشت که اصطلاح داده کاوی زمانی به کار برده می شود که با حجم بزرگی از داده ها، در حد مگابایت تا ترابایت، مواجه باشیم. در تمامی منابع داده کاوی، بر این مطلب تاکید شده است. هر چه حجم داده ها بیشتر و روابط میان آنها پیچیده تر باشد، دسترسی به اطلاعات نهفته در میان داده ها مشکل تر میشود و نقش داده کاوی به عنوان یکی از روشهای کشف دانش، روشن تر می گردد.