

۱۳۹۶۳۱۲

داده‌کاوی ۲

دکتر جمال شهباز

عضو هیأت علمی دانشگاه صنعتی امیرکبیر

و دبیر کنفرانس‌های داده‌کاوی ایران

انتشارات سروش گیتا

سرشناسه	: شهرابی، جمال، ۱۳۴۵ -
عنوان و نام پدیدآور	: داده کاوی ۱/۲ جمال شهرابی.
وضعیت ویراست	: ویراست ۳.
مشخصات نشر	: تهران: سروش گیتا، ۱۳۹۴.
مشخصات ظاهری	: ۲۹۹ص: مصور، جدول، نمودار.
شابک	: ۲۳۰۰۰۰ ریال ۵-۲-۹۵۶۱۴-۶۰۰-۹۷۸ :
وضعیت فهرست نویسی	: فیفا
یادداشت	: چاپ قبلی: جهاد دانشگاهی، واحد صنعتی امیرکبیر، ۱۳۹۰.
یادداشت	: واژه نامه.
موضوع	: داده کاوی
رده بندی شنگره	: ۹/۷۶QA ۱۳۹۴ ۹۴ش/د۲
رده بندی دیویی	: ۰۰۶/۳۱۲
شماره کتابشناسی ملی	: ۴۱۱۶۶۸۷



داده کاوی ۱

دکتر جمال شهرابی

چاپ سوم، ۱۳۹۴، شمارگان: ۱۵۰۰ نسخه، چاپ: قلم، قیمت: ۲۳۰۰۰۰ ریال
تمام حقوق این اثر برای انتشارات سروش گیتا محفوظ است.

خیابان اقدسیه، بزرگراه ارتش، مجتمع میلاد، بلوک ۳، واحد ۱۶، کد پستی: ۱۶۹۵۸ - ۳۵۴۴۱

تلفن: ۲۲۴۹۷۲۸۳ و ۲۲۴۹۷۲۹۳ - تلفکس: ۲۲۴۹۶۶۹۳

Publication@irandatamining.com

www.irandatamining.com

پیشگفتار انتشارات سروش گیتا:

دانش نوین داده‌کاوی، هوشمندی کسب و کار و هوش مصنوعی زمینه‌های تخصصی انتشارات سروش گیتا است. این انتشارات با انتشار اولین و بیشترین کتب داده‌کاوی در کشور مقام اول را در گسترش دانش داده‌کاوی دارا است.

انتشارات سروش گیتا تلاش دارد با همکاری تمامی متخصصان و علاقه‌مندان دانش داده‌کاوی و هوشمندی کسب و کار و هوش مصنوعی، در انتشار مستمر آخرین یافته‌های علمی در این زمینه، نقشی ارزشمند در گسترش این زمینه‌های تخصصی در کشور داشته باشد.

انتشارات سروش گیتا از تمامی زحمات و مشارکت‌های این عرصه‌ها جهت چاپ و انتشار آثار ارزشمندشان دعوت به عمل می‌آورد.

پیشگفتار:

دانش نوین داده‌کاوی (Data Mining) یکی از ده دانش در حال توسعه‌ای است که دهه آینده را با انقلاب تکنولوژیک مواجه خواهد ساخت و بدین رو در سال‌های اخیر در دنیا گسترش فوق‌العاده سریعی داشته است. دانش داده‌کاوی فرایند کشف دانش پنهان درون داده‌ها است که با برخورداری از دامنه وسیع زیرزمینه‌های تخصصی با توصیف، تشریح، پیش‌بینی و کنترل پدیده‌های گوناگون پیرامونی، امروزه دارای کاربرد بسیار وسیع در حوزه‌های مختلف است به گونه‌ای که امروزه مرز و محدودیتی برای کاربرد این دانش در نظر گرفته نشده و زمینه‌های کاری این دانش را از ذرات کف اقیانوس تا اعماق فضا می‌دانند.

در طول دهه گذشته حجم زیادی از داده‌ها در پایگاه داده‌ها انباشته و ذخیره شده‌اند و نتیجه این انباشتگی این است که سازمان‌ها در داده غنی ولی در کسب دانش بسیار ضعیف می‌باشند. امروزه میزان داده‌های در دسترس هر کارگزار دو برابر می‌شود و سازمانی توانا است که قادر باشد حداقل ۷ درصد از اطلاعاتش را مدیریت کنند. تحقیقات انجام یافته نشان از آن دارد که سازمان‌ها امروزه کمتر از یک درصد از داده‌هایشان را برای تحلیل استفاده می‌نمایند. به عبارت دیگر امروزه سازمان‌ها در اطلاعات غرق شده‌اند در حالی که برای بهره‌گیری از دانش هستند. چرا که سازمان‌ها داده‌های زیادی را در تصرف خود دارند در حالی که هنوز با فقدان دانش پیرامون این داده‌ها مواجه هستند.

با توجه به تنوع زیاد مخاطبین، مشتریان، بازارها، تنوع و پیچیدگی خدمات و محیط‌های کسب و کار، دسترسی به اطلاعات مناسب برای تصمیم‌گیری صحیح ضروری می‌باشد. از این رو استفاده از راهکارهای مناسب برای طبقه‌بندی و تولید اطلاعات از میان انبوهی از داده‌ها برای سازمان‌ها امری ضروری و حیاتی است. علم داده‌کاوی پاسخی به این نیاز است. منظور از کشف دانش پنهان داده‌ها و تأمین اطلاعات مورد نیاز می‌باشد. دانش داده‌کاوی ابزاری است که تصمیم‌گیرندگان را قادر می‌سازد تا سریع‌تر نسبت به آینده عمل نمایند، فعال باشند به جای آن که واکنش‌پذیر باشند، بدانند و مطمئن باشند به جای آن که حدس بزنند. واضح است که با داشتن این قابلیت سازمان‌ها قادر می‌باشند دانش ارزشمندی را از داده‌هایشان بدست آورند.

فن‌آوری داده‌کاوی سازمان‌ها را قادر می‌سازد تا از سرمایه داده‌هایشان بهره‌گیری نمایند. این ابزار برای پشتیبانی فرایند تصمیم‌گیری استفاده می‌گردد. داده‌کاوی پردازش بهینه تصمیم‌گیری را در سازمان‌ها تسهیل نموده و با انجام فرایند تصمیم‌سازی از طریق استخراج دانش با ارزش از داده، تصمیم‌گیری را برای مدیران سازمان تسهیل می‌نماید. 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.

با توجه به آن که داده‌کاوی فرآیندی است که از انواع تکنیک‌های مدل‌سازی و آنالیز داده برای کشف الگوها و ارتباطات در داده‌ها که ممکن است برای انجام پیش‌بینی‌های دقیق مورد استفاده قرار گیرد بهره می‌برد، این امر سازمان‌ها را در موضوعات متنوعی یاری می‌دهد. امروزه حجم انبوه اطلاعات

و تعاملات پیچیده و فزاینده، داده‌کاوی را به صف مقدم فرایند سودآور نمودن سازمان‌ها مبدل نموده است. به نظر می‌رسد در جامعه نوین کنونی سازمان‌ها و موسسات، استفاده از روشهای سنتی جمع‌آوری و تحلیل داده به دلیل اتلاف زمان و ایجاد هزینه‌های بسیار زیاد و عدم استفاده از سیستم‌های تصمیم‌گیری هوشمند مدرن مناسب نبوده و از این رو استفاده از روشهای جدید آنالیز داده مانند داده‌کاوی برای پایگاه‌های داده این صنعت بسیار حیاتی به نظر می‌رسد.

رقابتی شدن بازارهای جهانی و منطقه‌ای، موسسات را ملزم به ایجاد مزیت‌های رقابتی در تولید، ارائه خدمات و افزایش رضایتمندی و جذب بیشتر مشتریان نموده است تا از این طریق بتوانند جایگاهی مناسب‌تر میان مشتریان کسب کرده و بر رشد سازمان بیفزایند.

با توجه به جایگاه دانش داده‌کاوی کتاب حاضر تلاش دارد گامی در جهت گسترش آن در کشور بردارد. جلد نخست این اثر همزمان با برپایی موفق اولین کنفرانس داده‌کاوی در سال ۱۳۸۶ منتشر گردید. با استقبال گسترده مخاطبین، چاپ دوم آن با ویرایش کلی، همزمان با این کتاب در آستانه پنجمین تهران داده‌کاوی ایران منتشر گردیده است. این کتاب، تدوین و برداشتی آزاد از منابع مختلف از جمله کتاب داده‌کاوی آقای مایکل بری می‌باشد که با بومی‌سازی، به روز رسانی و اضافه نمودن توضیحات مختلف درک موضوع را برای مخاطبین ساده و هموار می‌سازد. این کتاب به معرفی و توضیح تکنیک‌های داده‌کاوی پرداخته است. تشریح فنی و تخصصی تکنیک‌های داده‌کاوی قبلاً در کتاب داده‌کاوی پیشرفته، توسط اینجانب منتشر گردیده است.

در اینجا بر خود لازم می‌دانم نصیب و عنایت خاص حضرت حق که بدون هدایت و راهنمایی ملوکانه حضرت ایشان توفیق انجام این مهم برایم فراهم داشت را حمد و سپاس گویم. همچنین از تمامی عزیزانی که مرا در این مسیر پشتیبان بودند، از پدر و مادر عزیزم که پس از خداوند باری تعالی شکرگزاری آنان بر من واجب است، از همسر گرانقدرم که در تمامی ارتباط، حمایت و تشویق و یاری ایشان همواره همراه من بوده است تشکر و قدردانی می‌نمایم.

بر خود لازم می‌دانم از تلاش‌های سرکار خانم مهندس ابدسم زرگرچی که مرا در انجام این مهم یاری رساندند، کمال تشکر را نمایم. همچنین از خانم حوریه کفاش تهرانی و آقای احمدرضا مختاری مدیر محترم انتشارات جهاد دانشگاهی صنعتی امیرکبیر سپاسگزاری می‌نمایم.

جمال شهرابی

عضو هیات علمی دانشکده مهندسی صنایع

دانشگاه صنعتی امیرکبیر

و دبیر کنفرانس داده‌کاوی ایران

Jamalshahrabi@aut.ac.ir