

داده‌کاوی ۲

تألیف

دکتر جمال شهرابی

عضو هیأت علمی دانشکده مهندسی صنایع و سیستم‌های مدیریت

دانشگاه صنعتی امیرکبیر

و دبیر کنفرانس داده‌کاوی ایران

انتشارات جهاد دانشگاهی

واحد صنعتی امیرکبیر

سرشناسه: شهرابی، جمال- ۱۳۴۵
 عنوان و نام پدیدآور: داده‌کاوی ۲ / تألیف جمال شهرابی
 مشخصات نشر: تهران: جهاد دانشگاهی، واحد صنعتی امیرکبیر، ۱۳۹۰.
 مشخصات ظاهری: ۳۶۹ ص: ۸ مصور
 شابک: 978-964-210-009-5 : ۱۰۰/۰۰۰ ریال،
 وضعیت فهرست‌نویسی: فیبا
 یادداشت: کتابنامه ۳۶۹ ص
 موضوع: درخت‌های تصمیم
 موضوع: شبکه‌های عصبی مصنوعی
 موضوع: اندازگیری فاصله‌ها
 موضوع: خوش‌بندی
 موضوع: الگوریتم و ژنتیک
 شناسه افزوده: جهاد دانشگاهی، واحد صنعتی امیرکبیر
 رده‌بندی کنگره: ۱۳۹۰ ۹۴۲/ش ۴۱۵/۵ GF
 رده‌بندی دیویی: ۶۵۸۸۱۲
 شماره کتابشناسی ملی: ۱۱۲۳۳۴۷

انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر

• نام کتاب	• داده‌کاوی ۲
• تألیف	• دکتر جمال شهرابی
• نوبت چاپ	• اول
• سال چاپ	• ۱۳۹۰
• شمارگان	• ۱۰۰۰ نسخه
• قیمت	• ۱۰۰۰۰۰ ریال
• چاپخانه	• قلم
• شابک	• 978-964-210-009-5 ISBN
• مرکز پخش	• ۹۷۸-۹۶۴-۲۱۰-۰۰۹-۵

دیرپرخانه دائمی کتفرانس داده‌کلوی ایران

irandatamining@gmail.com

www.irandatamining.com



پیشگفتار جهاد دانشگاهی

توجه به تقویت بنیه علمی دانشگاه‌ها و دیگر مراکز علمی یکی از وظایف اصلی همه آگاهان و دست‌اندرکاران این امور در دوران بازسازی کشور اسلامی عزیزمان ایران می‌باشد. در همین راستا جهاد دانشگاهی با آگاهی نسبت به این موضوع، تمامی توان خود را صرف افزایش سطح علمی دانشجویان و آینده‌سازان کشورمان نموده است و در این مسیر از هیچ کوشش و تلاشی دریغ نمی‌کند.

انتشارات جهاد دانشگاهی با این اعتقاد که دستیابی به استقلال علمی، اقتصادی و عقیدتی بدون دستیابی اولیه به سطوح بالای علمی و تکنولوژی‌های مدرن در کنار ایمان و اعتقاد، امکان‌پذیر نمی‌باشد، سعی می‌کند با آماده‌سازی و انتشار متون متنوع علمی به هر دو زبان فارسی و لاتین که از سطح مطلوب علمی نیز برخوردار باشند به وظیفه انسانی و اعتقادی خویش عمل کند و امیدوار است که بتواند کمکی هرچند ناچیز به علاقه‌مندان و دلسوزان میهن اسلامی و پویندگان راه علم انجام دهد.

با این باور و اعتقاد از همه عزیزان دانش‌پژوه جهت ارایه نظرات و پیشنهادات برای هرچه بهتر شدن سطح علمی انتشارات این جهاد، یاری می‌طلبیم.

در انتها، این کتاب را به روان پاک شهدای دانشگاهی بویژه شهدای دانشگاه صنعتی امیرکبیر تقدیم می‌نماییم. امید است کلیه دانشگاهیان استوار و ثابت‌قدم، پوینده راه شهیدان در محیط مقدس دانشگاه باشند.

والسلام علی من التبع الهدی

انتشارات جهاد دانشگاهی

واحد صنعتی امیرکبیر

مدیر مسئول

احمد رضا مختاری

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 ۳ سال دو برابر می‌شود و سازمانی توانا است که قادر باشد حداقل ۷ درصد از اطلاعاتش را مدیریت نمایند. 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 گرسنه دانش هستند چرا که سازمان‌ها داده‌های زیادی را در تصرف خود دارند در حالی که هنوز با فقدان دانش پنهان درون 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 به منظور کشف دانش پنهان داده‌ها و تامین 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 و 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 بهره‌برداری 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 و با استقبال گسترده مخاطبین، چاپ دوم آن با ویرایش کلی، همزمان با این کتاب در آستانه پنجمین کنفرانس 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 لطف و عنایت خاص حضرت حق که بدون هدایت و راهنمایی ملوکانه حضرت ایشان توفیق انجام این مهم برایم فراهم نمی‌گشت را حمد و سپاس گویم. همچنین از تمامی عزیزانی که مرا در این مسیر پشتیبان بودند، از پدر و مادر عزیزم که پس از خواندن باریتمالی شکرگزاری آنان بر من واجب است، از همسر گرانقدرم که در تمامی شرایط حمایت و تشویق و یاری ایشان همواره همراه من بوده است تشکر و قدردانی می‌نمایم.

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

جمال شهبابی

عضو هیات علمی دانشکده مهندسی صنایع

دانشگاه صنعتی امیرکبیر

و دبیر کنفرانس داده‌کاوی ایران

Jamalshahrabi@aut.ac.ir

فهرست مطالب

۱	۱. درخت‌های تصمیم
۵۷	۲. شبکه‌های عصبی مصنوعی
۱۲۳	۳. تکنیک‌های نزدیک‌ترین همسایه
۱۶۱	۴. تحلیل سبد خرید و قوانین وابستگی
۱۹۹	۵. تحلیل پیوند
۲۳۱	۶. خوشه‌بندی
۲۶۵	۷. الگوریتم ژنتیک