

داده‌کاوی با کلمنتاین

دکتر جمال شهرابی

عضو هیئت علمی دانشکده مهندسی صنایع و سیستم‌های مدیریت

دانشگاه صنعتی امیرکبیر

و دبیر کنفرانس داده‌کاوی ایران

مهندس ابوالفضل زارع

کارشناس ارشد مهندسی صنایع

دانشگاه صنعتی امیرکبیر

مرشنامه : شهرابی، جمال، ۱۳۴۵ -
 عنوان و نام پدیدآور : داده‌کاوی با کلمتاین / تألیف جمال شهرابی، ابوالفضل زارع.
 مشخصات نشر : تهران: جهاد دانشگاهی، واحد صنعتی امیرکبیر، ۱۳۹۲.
 مشخصات ظاهری : ۲۸۰ ص.: مصور، نمودار.
 شابک : ۲۰۰۰۰۰ ریال 8-165-210-964-978:
 وضعیت فهرست نویسی : فیبا
 موضوع : نرم‌افزار کلمتاین
 موضوع : داده‌کاوی --- نرم‌افزار
 موضوع : داده‌کاوی
 شناسه افزوده : زارع، ابوالفضل، ۱۳۶۴ -
 شناسه افزوده : جهاد دانشگاهی. واحد صنعتی امیرکبیر
 رده بندی کنگره : QA۷۷۹ ۱۳۹۲ ۹۲ش ۵۲/
 رده بندی دیویی : ۰۰۵/۷۴
 شماره کتابشناسی ملی : ۳۴۳۰۶۳



انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر

- نام کتاب : داده‌کاوی با کلمتاین
- تألیف : دکتر جمال شهرابی - مهندس ابوالفضل زارع
- نوبت چاپ : اول
- سال چاپ : ۱۳۹۲
- شمارگان : ۱۰۰۰ نسخه
- قیمت : ۲۰۰۰۰۰ ریال
- چاپخانه : قلم
- شابک : ۹۷۸-۹۶۴-۲۱۰-۱۶۵-۸ : ISBN 978-964-210-165-8
- مرکز پخش: دبیرخانه دائمی کنفرانس داده‌کاوی ایران،

۲۲۴۹۷۲۸۳ ، ۲۲۴۹۷۲۹۳ ، www.Iran www.Irandatamining.com

تمامی حقوق این اثر برای موسسه پژوهشی داده‌کاوی گیتا محفوظ است.



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 آمادگی 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 هر چند ناچیز به علاقه‌مندان و دلسوزان میهن اسلامی و پیوندگان راه علم انجام دهد.

با این باور و اعتقاد از همه عزیزان دانش‌پژوه جهت ارایه نظرات و پیشنهادات برای هر چه 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، پیونده راه شهیدان در محیط مقدس دانشگاه باشند.

والسلام علی من‌التبع الهدی

انتشارات جهاد دانشگاهی

واحد صنعتی امیرکبیر

مدیر مسئول

احمد رضا مختاری

تقدیم به تمام کسانی که دل در گرو وطن دارند،

تقدیم به آنانی که آینده میهن و اعتلای روزافزونی دغدغه آنهاست،

تقدیم به افرادی که بهروزی مردم آرزوی ایشان است و

تقدیم به کسانی که بهره‌گیری از دانش‌های نوین را تنها راه اعتلای میهن

و بهروزی ملت می‌دانند و در این راه تلاش می‌کنند.

پیشگفتار:

دانش نوین داده‌کاوی (Data Mining) دانش در حال توسعه‌ای است که دهه حاضر را با انقلاب تکنولوژیک مواجه ساخته و بدین رو در سال‌های اخیر در دنیا گسترش فوق‌العاده سریعی داشته است. دانش داده‌کاوی فرآیند 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. امروزه داده‌کاوی مدیران سازمان‌ها را در موضوعات متنوعی یاری داده بگونه‌ای که استفاده از دانش نوین داده‌کاوی به عنوان جدیدترین و برترین روش حل مسئله برای پایگاه‌های داده سازمان‌ها بسیار حیاتی بوده و بدین جهت امروزه این دانش در تمامی موضوعات و برنامه‌ریزی‌های خرد و کلان، نقشی موثر ایفا می‌نماید.

دانش داده‌کاوی همزمان با ایجاد چندین فرصت و ویژگی در دنیای کسب و کار در دهه اخیر در دنیا متولد گردیده است، از یک سو امکان جمع‌آوری داده‌ها در مقیاس بزرگ برای سازمان‌ها فراهم گردیده، از سوی دیگر امکان انجام محاسبات پیچیده با افزایش توان تحلیلی رایانه‌ها ایجاد گردیده، همچنین الگوریتم‌های پیشرفته خلق گردیده و از همه مهمتر پیچیدگی‌های کسب و کار رقابتی بیش از گذشته شده و نیاز مدیران سازمان‌ها به داشتن مزیت رقابتی فزونی یافته و دراین شرایط دانش داده‌کاوی به عنوان منجی سازمان‌های برخوردار از داده، با ایجاد مدل‌های مبتنی بر داده و ساختن سیستم‌های تصمیم‌یار، امکان اشراف کامل، جامع، صحیح، به موقع و پویای مدیران را بر سازمان خرد و رقبا و محیط کسب و کار فراهم مینماید.

امروزه جایی و موضوعی در دنیای پیشرفته یافت نمیشود که برخوردار از دانش داده‌کاوی نباشد. بر حسب اعلام دانشگاه MIT به عنوان پیشرو دانش‌های نوین، "امروزه مرز دانش داده‌کاوی از اعماق اقیانوس‌ها تا بیکران فضا است" و در تمامی زمینه‌های برخوردار از داده کاربرد فراوان دارد. امروزه دنیا با استفاده از دانش نوین داده‌کاوی به عنوان جدیدترین و برترین روش حل مسئله، تمامی موضوعات مبتلا به خود که قبلاً با روش‌های سنتی قابل حل نبودند را تحلیل نموده و دستاوردهای ارزشمندی به ارمغان آورده است.

دبیرخانه دائمی کنفرانس داده‌کاوی ایران به عنوان "پدر داده‌کاوی ایران" از سال ۱۳۸۶ با معرفی دانش داده‌کاوی موجب گسترش این دانش در کشور بوده و جهت تحقق این امر

گام‌های بلندی برداشته است که برگزاری کنفرانس‌های سالانه داده کاوی ایران و انتشار کتب تخصصی داده کاوی از آن جمله است. کتاب حاضر، یازدهمین کتاب تخصصی داده کاوی منتشر شده توسط این دبیرخانه است که کوشیده است راهنمایی کامل و ارزشمند و برخوردار از نوشتاری روان برای آموزش نرم‌افزار کلمنتاین برای دوستداران بهره‌گیری از دانش داده کاوی باشد. مثال‌های واقعی و گام به گام توضیح داده شده در فصل آخر این کتاب بهره‌مندی عملی از این دانش را برای خواننده آسان و روان می‌سازد. این کتاب برای علاقه‌مندان بکارگیری دانش داده کاوی به ویژه دانشجویان تحصیلات تکمیلی کلیه رشته‌ها به خصوص مهندسی صنایع، مهندسی کامپیوتر، و ریاضی تدوین گردیده است. پدیدآورندگان این اثر آماده دریافت نظرات و پیشنهادات، و انتقادات خوانندگان گرامی از طریق آدرس پست الکترونیکی info@irandatamining.com می‌باشند.

دبیرخانه دائمی کنفرانس داده کاوی ایران

فهرست مطالب

فصل یک - داده و داده کاوی	۱
فصل دو - متدولوژی کریسپ	۲۷
فصل سه - الگوریتم‌ها	۵۷
فصل چهار - کار با SPSS Modeler	۸۳
فصل پنج - گره‌ها	۱۱۱
فصل شش - مثال‌های کاربردی با کلمنتاین	۱۶۷