

داده کاوی با کلمنتاین

تالیف

دکتر جمال شهرابی

عضو هیئت علمی دانشکده مهندسی صنایع و سیستم های مدیریت

دانشگاه صنعتی امیرکبیر

و دبیر کنفرانس داده کاوی ایران

مهندس ابوالفضل زارع

انتشارات سروش گیتا

سرشناسه	: شهرابی، جمال، ۱۳۴۵ -
عنوان و نام پدیدآور	: داده‌کاوی در کلمنتاین / تألیف جمال شهرابی، ابوالفضل زارع.
مشخصات نشر	: تهران: سروش گیتا، ۱۳۹۴.
مشخصات ظاهری	: ۱۶۰ ص.: مصور، جدول، نمودار.
شابک	: ۹۷۸-۶۰۰-۹۶۲۹۰-۱-۵
وضعیت فهرست نویسی	: فیپا
یادداشت	: چاپ قبلی: جهاد دانشگاهی، واحد صنعتی امیرکبیر، ۱۳۹۲ ([۶]، ۲۶۸ ص.).
یادداشت	: واژه نامه
موضوع	: نرم‌افزار آی. بی. ام. اس. پی. اس. اس. مودلر
موضوع	: داده‌کاوی -- نرم‌افزار
موضوع	: داده‌کاوی
شماره افزوده	: زارع، ابوالفضل، ۱۳۶۴ -
رده بندی کنگ	: ۹/۷۶QA ۱۳۹۴ ۹۲ش۵۲/
رده بندی دیویی	: ۰۰۵/۷۴
شماره کتابشناسی	: ۴۱۵۹۰



داده کاوی با کلمنتاین

دکتر جمال شهرابی

مهندس ابوالفضل زارع

چاپ اول، ۱۳۹۴، شمارگان: ۱۰۰۰ نسخه، چاپ: قلم، قیمت: ۲۰۰۰۰۰ 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 زمینه‌ای تخصصی در کشور داشته باشد.

انتشارات سروش گیتا ز تمامی فرهیختگان این عرصه‌ها جهت چاپ و انتشار آثار ارزشمندشان دعوت به عمل می‌آورد.

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 مدبران سازمان‌ها را در موضوعات متنوعی یاری داده بگونه‌ای که استفاده از دانش نوین داده‌کاوی به آنان جدیدترین و برترین روش حل مسئله برای پایگاه‌های داده سازمان‌ها بسیار حیاتی بود و بدین جهت امروزه این دانش در تمامی موضوعات و برنامه‌ریزی‌های خرد و کلان، نقش مهمی ایفا می‌نماید.

دانش داده‌کاوی همزمان با ایجاد 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 خود 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 موجد 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 الکترونیک 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 المنتاین	۱۶۷