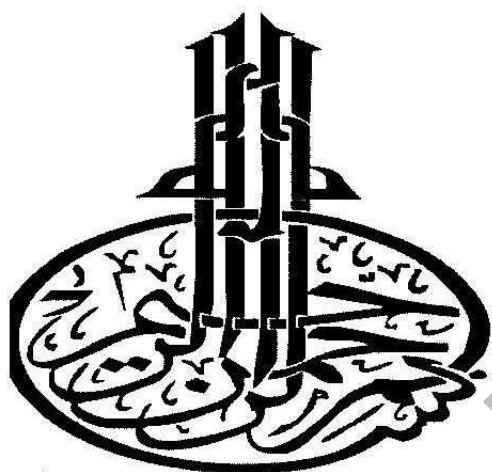




امتحان اولیه داده کاوی

تألیف

سهیلا ملکی

کارشناس ارشد مهندسی فناوری اطلاعات



سرشناسه : ملکی، سهیلا، ۱۳۶۵ فروردین -
عنوان و نام پدیدآور : اصول اولیه داده کاوی / تألیف سهیلا ملکی.
مشخصات نشر : ایلام، مکتب خانه، ۱۳۹۹.
مشخصات ظاهری : ۱۰۳ ص.: مصور، جدول و نمودار.
شابک : ۹۷۸-۶۲۲-۹۵۱۳۳-۶-۱ :
وضعیت فهرست نویسی: فیپا
یادداشت : کتابنامه: ص. ۱۰۲.
موضوع : داده کاوی
موضوع : Data mining
رده بندی کنگره : Q۱۶۱
رده بندی دیویی : ۰۶۲۱۲
شماره کتابشناسی ملی : ۱۹۲۸۳۹

عنوان: اصول اولیه داده کاوی

مؤلف: سهیلا ملکی

ویراستار: وحیده میرزائی

ویرایش گر: ساسان تیموری

چاپ اول: ۱۳۹۹

شمارگان: ۵۰۰

قیمت: ۲۰۰۰۰ تومان

فهرست مطالب

عنوان	صفحه
پیشگفتار	۱۰
فصل اول: کلیات	
۱-۱- مقدمه	۱۲
۲-۱- حرکت به سوی عصر اطلاعات	۱۲
۳-۱- داده ناوی به عنوان سیر تکاملی فن آوری اطلاعات	۱۳
۴-۱- داده کاوی چیست؟	۱۸
۵-۱- چه نوعی از داده ها کاوش می شوند؟	۲۰
فصل دوم: اصول بنیادی داده ها	
۱-۲- داده های پایگاه داده ها	۲۲
۲-۲- انبارهای داده	۲۳
۳-۲- داده های تراکنشی	۲۴
۴-۲- گونه های دیگر داده ها	۲۵
۵-۲- چه نوعی از الگوها کاوش می شوند؟	۲۶
۱-۵-۲- توصیف و تفکیک	۲۶
۲-۵-۲- کاوش الگوهای مکرر، مشارکت ها و وابستگی ها	۲۸
۳-۵-۲- دسته بندی و رگرسیون برای تحلیل پیش گویی	۲۸
۴-۵-۲- تحلیل خوشه	۳۰
۵-۵-۲- تحلیل داده های پرت	۳۰

- ۳۱ ۲-۶- آیا کلیه ی الگوها جالب هستند؟
- ۳۳ ۲-۷- کدام فن آوری ها استفاده می شوند؟
- ۳۴ ۲-۷-۱- آمار
- ۳۵ ۲-۷-۲- یادگیری ماشین
- ۳۶ ۲-۸- هدف داده کاوی کدام نوع از برنامه های کاربردی است؟
- ۴۴ ۲-۹- شناخت داده ها
- ۴۴ ۲-۹-۱- گونه ها و انواع صفات خاصه

فصل سوم: عملیات قبل از پردازش

- ۴۸ ۳-۱- آمار توصیفی
- ۴۸ ۳-۱-۱- محاسبه ی شاخص های مرکزی: میانگین، میانه و مد
- ۴۸ ۳-۱-۲- محاسبه ی شاخص های پراکندگی داده ها
- ۴۹ ۳-۱-۳- نمایش گرافیکی آمار توصیفی
- ۴۹ ۳-۱-۳-۱- رسم نمودار چندک
- ۵۰ ۳-۱-۳-۲- نمودار چندک - چندک
- ۵۱ ۳-۱-۳-۳- هیستوگرام ها
- ۵۲ ۳-۱-۳-۴- نمودارهای پراکنشی و همبستگی داده ها
- ۵۳ ۳-۲- مصور سازی داده ها
- ۵۴ ۳-۳- روش های مصور سازی پیکسل گرا
- ۵۵ ۳-۴- تکنیک های تصویر کردن هندسی
- ۵۷ ۳-۵- تکنیک های مصور سازی مبتنی بر شمایل

- ۳-۶- تکنیک های سلسله مراتبی مصورسازی ۵۸
- ۳-۷- مصورسازی داده ها و روابط پیچیده ۶۰
- ۳-۸- محاسبه ی تشابه و عدم تشابه ۶۱
- ۳-۹- محاسبه ی همسایگی برای صفات خاصه ی اسمی ۶۲
- ۳-۱۰- محاسبه ی همسایگی برای صفات خاصه ی دودویی ۶۲
- ۳-۱۱- پیش پردازش داده ها ۶۳
- ۳-۱۱-۱- مروری بر پیش پردازش داده ها ۶۳
- ۳-۱۱-۲- وظایف اصلی در پیش پردازش داده ها ۶۴
- ۳-۱۲- پالایش داده ها ۶۶
- ۳-۱۳- یکپارچه سازی داده ها ۷۰
- ۳-۱۴- کاهش داده ها ۷۳
- ۳-۱۵- تبدیل داده ها و گسسته سازی داده ها ۸۳
- فصل چهارم: انبارش داده ها و تحلیل بر خط**
- ۴-۱- انبار داده ها: مفاهیم پایه ای ۸۷
- ۴-۱-۱- انبار داده ها چیست؟ ۸۷
- ۴-۲- تفاوت های میان سیستم های پایگاه داده عملیاتی و انباری داده ها ۸۸
- ۴-۳- انبارش داده ها: یک معماری چند ردیفی ۹۱
- ۴-۴- مدل های انبار داده ها ۹۳
- ۴-۵- سیستم های OLAP در مقابل پایگاه داده های آماری ۱۰۰
- مراجع ۱۰۲

پیشگفتار

سعی بر این بود که مطالب ساده وبا بیانی عامیانه برای خوانندگان توضیح داده شود، زیرا داده کاوی فقط مختص علوم کامپیوتر نیست و بسیار در پژوهش های میان رشته ای کاربرد دارد. به همین دلیل درکی درست از داده کاوی برای پژوهشگران بسیار واجب است. هر پژوهشی دارای نواقصی است اما تلاش ما با کمک گرفتن از منابع علمی برای کم کردن این نواقص بود. در دنیای امروزه که اقتصاد جهانی شده شرکت ها و سازمان شدت رقابت را به درستی درک کرده اند. زمانی رقابت می تواند برای یک شرکت باعث پیشرفت شود که اطلاعات بعنوان یکی از فاکتورهای تولیدی مهم پدیدار شده باشد. بر همین سبب تلاش برای استخراج و جمع آوری اطلاعات توجه بسیاری از صنایع را به خود جلب کرده است. این که داده چیست، چگونه جمع آوری می شود و نحوه دسته بندی آنها می باشد، در کاوش داده ها بسیار مفید باشد. داده کاوی به مفهوم استخراج اطلاعات نهان یا الگوها و رابطه مشخص در حجم زیادی از داده ها در یک یا چند بانک اطلاعاتی بزرگ گفته می شود. داده کاوی، پایگاه ها و مجموعه حجیم داده ها را در پی کشف و استخراج، مورد تحلیل قرار می دهد. این گونه مطالعات و کاوش ها را به واقع می توان همان امتداد و استمرار دانش کهن و همه جا گیر آمار دانست. تفاوت عمده در مقیاس، وسعت، گوناگونی زمینه ها، کاردها و نیز ابعاد و اندازه های داده های امروزی است که شیوه های ماشینی مربوط به بازیگری، مدل سازی، و آموزش را طلب می نماید.