

داده کاوی با Weka

مؤلفین:

مهندس مهران خسروی - مهندس پیمان بهرام پور

سرشناسه	:	خسروی، مهران، ۱۳۶۶ -
عنوان و نام پدیدآور	:	داده کاوی با weka / مولف مهران خسروی، پیمان بهرام پور.
مشخصات نشر	:	تهران: آرنا، ۱۳۹۳.
مشخصات ظاهری	:	۲۵۲ ص.
شابک	:	978-600-356-040-6
وضعیت فهرست نویسی	:	فیا
موضوع	:	داده کاوی -- نرم افزار
موضوع	:	الگوریتم های کامپیوتری
شناسه افز	:	بهرام پور، پیمان، ۱۳۶۴ -
رده بندی کنگره	:	۹/۷۶QA / د ۱۳۹۳۲۵ خ
رده بندی دیویی	:	۳۱۲/۰۰۶
شماره کتابشناسی ملی	:	۳۴۹۱۱۰۳



نشر آرنا

عنوان	:	داده کاوی با weka
مولفین	:	مهران خسروی، پیمان بهرام پور
ناشر	:	آرنا
صفحه آرایي	:	الهه بذرافشان
چاپ	:	اول ۱۳۹۳
شمارگان	:	۱۰۰۰ نسخه قیمت: ۱۴۰۰۰ تومان
شابک	:	۹۷۸-۶۰۰-۳۵۶-۰۴۰-۶

www.arnapub.com

فهرست مطالب

۱۱	پیشگفتار
۱۳	فصل اول: تعاریف و مفاهیم
۱۵	مفاهیم پایه در داده کاوی
۱۵	داده، اطلاعات و دانش
۱۶	تاریخچه داده کاوی
۱۷	تعریف داده کاوی
۲۰	مراحل فرایند کشف دانش از پایگاه داده ها
۲۳	عملیاتهای داده کاوی
۲۴	مدلسازی پیشگویی کننده
۲۴	تقطیع پایگاه داده ها
۲۵	تحلیل پیوند
۲۷	معماری یک سیستم داده کاوی
۲۸	قابلیتهای داده کاوی
۲۹	داده کاوی و انبار داده ها
۳۰	داده کاوی و OLAP
۳۰	داده کاوی، آمار و یادگیری ماشین
۳۱	کاربردهای داده کاوی
۳۲	داده کاوی موفق
۳۲	تحلیل ارتباطات
۳۶	سلسله مراتبی از انتخاب ها
۴۰	مدلها و الگوریتمهای داده کاوی

۴۰	شبکه های عصبی
۴۱	درخت های انتخاب
۴۲	استنتاج قانون
۴۲	الگوریتمهای ژنتیک
۴۳	فرآیند داده کاوی
۴۳	مدلهای فازی
۴۴	مدل فرایند سوپره
۵۱	فصل دوم: الگوریتم های دسته بندی
۵۲	مقدمه
۵۲	روش های آماری
۵۳	استنتاج آماری
۵۴	ارزیابی تفاوت ها در مجموعه داده ها
۵۵	استنتاج بیزین
۵۶	رگرسیون پیش گوینه
۶۴	تحلیل واریانس
۶۸	رگرسیون لجستیک
۷۰	مدل Log خطی
۷۶	تحلیل ممیز خطی
۷۸	درخت های تصمیم گیری
۷۹	انواع درختان تصمیم
۸۰	اهداف اصلی درخت های تصمیم گیری دسته بندی کننده
۸۰	جذابیت درختان تصمیم

۸۱	درختان رگرسیون
۸۲	تفاوت درخت رگرسیون و درخت دسته بند
۸۲	بازنمایی درخت تصمیم
۸۳	توسعه درختان تصمیم با گراف‌های تصمیم
۸۵	انواع یادگیری در درخت تصمیم گیری
۸۵	ویژگی‌های یک الگوریتم افزایشی خوب:
۸۶	مزایا و معایب درخت تصمیم
۸۷	معایب درختان تصمیم
۸۷	الگوریتم یادگیری درخت تصمیم بایه
۸۹	الگوریتم C4.5
۹۲	شبکه عصبی مصنوعی
۹۳	تاریخچه شبکه های عصبی مصنوعی
۹۵	چرا از شبکه های عصبی استفاده می کنیم
۹۶	مقایسه شبکه های عصبی با کامپیوترهای سنتی
۹۷	نورون مصنوعی
۹۷	ساختار شبکه های عصبی
۹۹	تقسیم بندی شبکه های عصبی
۱۰۰	کاربرد شبکه های عصبی
۱۰۱	معایب شبکه های عصبی
۱۰۱	الگوریتم‌های ژنتیک
۱۰۲	مقدمه‌ای بر بهینه سازی
۱۰۳	الگوریتم‌های مینیمم یابنده

هیوریستیک‌ها	۱۰۴
انواع الگوریتم‌های هیوریستیک	۱۰۶
معرفی کلی الگوریتم ژنتیک	۱۰۸
الگوریتم ژنتیک	۱۰۸
مکانیزم الگوریتم ژنتیک	۱۱۱
عملگرهای الگوریتم ژنتیک	۱۱۴
انواع الگوریتم‌های ژنتیکی	۱۱۵
الگوریتم ژنتیک سری	۱۱۶
الگوریتم ژنتیکی موازی	۱۱۷
برنامه ریزی ژنتیکی	۱۱۷
جمع بندی	۱۲۲
فصل سوم: خوشه بندی و قانون وابستگی	۱۲۳
تعریف خوشه بندی	۱۲۴
انواع خوشه ها	۱۲۶
الگوریتم های خوشه بندی	۱۲۷
خوشه بندی سلسله مراتبی (Hierarchical Clustering)	۱۲۷
الگوریتم خوشه بندی single-linkage	۱۲۸
الگوریتم خوشه بندی c میانگین (Fuzzy c-mean)	۱۳۵
K-means	۱۳۸
Mixture-Resolving and Mode-Seeking Algorithms	۱۴۱
EM Algorithm	۱۴۱
چه تحلیل و آنالیزهایی خوشه بندی نیستند؟	۱۴۲

۱۴۳.....	چالشهای الگوریتم های خوشه بندی.....
۱۴۴.....	قانون وابستگی.....
۱۴۵.....	یادگیری قانون وابستگی.....
۱۴۵.....	تعریف اصلی کاوش قانون وابستگی.....
۱۴۶.....	مفاهیم کاربردی.....
۱۴۷.....	الگوریتم.....
۱۵۱.....	فصل چهارم مقدمه ای بر Weka.....
۱۵۲.....	معرفی نرم افزار Weka.....
۱۵۵.....	روش استفاده از Weka.....
۱۵۶.....	تخمین پروژ های نرم افزار.....
۱۵۸.....	فاکتور های تخمین تلاش.....
۱۶۳.....	واسط Explorer سر برگ process.....
۱۷۸.....	واسط Explorer سر برگ Classify.....
۲۰۱.....	واسط Explorer سر برگ Cluster.....
۲۰۳.....	معیارهای تست خروجی الگوریتم.....
۲۱۴.....	واسط Explorer سر برگ Association.....
۲۱۹.....	واسط Explorer سر برگ Select Attributes.....
۲۲۷.....	واسط Explorer سر برگ Visualize.....
۲۲۸.....	واسط Experimenter.....
۲۳۹.....	واسط Knowledge Flow.....
۲۴۸.....	منابع.....

پیشگفتار

امروزه با گسترش سیستم های پایگاهی و حجم بالای داده های ذخیره شده در این سیستم ها، نیاز به ابزاری است تا بتوان داده های ذخیره شده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربران قرار داد. با استفاده از پرسش های ساده در SQL و ابزارهای گوناگون گزارش گیری معمولی، می توان اطلاعاتی را در اختیار کاربران قرار داد تا بتوانند به نتیجه گیری در مورد داده ها و روابط منطقی میان آنها بپردازند اما وقتی حجم داده ها بالا باشد، کاربران هر چند زیر دست و با تجربه باشند نمی توانند الگوهای مفید را در میان حجم انبوه داده ها تشخیص دهند و یا اگر قادر به این کار هم باشند، هزینه عدالت و نظر نیروی انسانی و مادی بسیار بالا است. از سوی دیگر کاربران معمولاً فرضیه ای را مطرح می کنند و سپس بر اساس گزارشات مشاهده شده به اثبات یا رد فرضیه می پردازند، در حالی که امروزه نیاز به روش هایی است که اصطلاحاً به کشف دانش بپردازند یعنی با کمترین دخالت کاربر و به صورت خودکار الگوها و رابطه های منطقی را بیان نمایند. داده کاوی یکی از بهترین این روش ها است که به وسیله آن الگوهای مفید در داده ها با حداقل دخالت کاربران شسته می شوند و اطلاعاتی را در اختیار کاربران و تحلیل گران قرار می دهند تا براساس آنها تصمیمات مهم و حیاتی در سازمان ها اتخاذ شوند. در داده کاوی از بخشی از علم کلامی به نام تحلیل اکتشافی داده ها استفاده می شود که در آن بر کشف اطلاعات نهفته و ناشناخته از روی حجم انبوه داده ها تاکید می شود. علاوه بر این داده کاوی با هوش مصنوعی و یادگیری ماشین نیز ارتباط تنگاتنگی دارد، بنابراین می توان گفت در داده کاوی تئوریهای پایگاه داده ها، هوش مصنوعی، یادگیری ماشین و علم آمار را در هم می آمیزند تا زمینه کاربردی فراهم شود. هر چه حجم داده ها بیشتر و روابط میان آنها پیچیده تر باشد دسترسی به اطلاعات نهفته در میان داده ها مشکل تر می شود و نقش داده کاوی به عنوان یکی از روشهای کشف دانش، روشن تر می گردد. کتاب حاضر با معرفی نرم افزار داده کاوی Weka به

عنوان ابزای توانمند برای تمامی سطوح کاربران تلاش کرده است تا کاربران را در انجام گام به گام تمامی مراحل داده کاوی از آماده سازی داده ها تا ارزیابی نتایج همراهی کند. به منظور آنکه کاربران نرم افزار Weka درک درستی از مفاهیم داده کاوی و الگوریتم های مختلف داشته باشند در فصل های ابتدایی این کتاب تلاش شده است هر چند به طور خلاصه اما بصورت جامع مفاهیم اولیه داده کاوی و الگوریتم های کاربردی در این زمینه معرفی و تشریح شوند تا کاربر با دید بهتری از نرم افزار Weka استفاده کند. تمامی بخش های نرم افزار Weka در غالب عکس ها و توضیحاتی کامل معرفی و نمونه بکارگیری آنها نشان داده شده است. امید است که کتاب حاضر امکانی را برای تمامی کاربران فراهم کند تا با هر سطحی از دانش داده کاوی بتواند فرآیندهای داده کاوی خود را در غالب نرم افزار Weka پیاده سازی و اجرا نمایند.

ومن ... توفیق