



به نام خدا

آموزش داده کاوی با Rapidminer

مؤلف:

دکتر ساسان کرمی زاده



هرگونه چاپ و تکثیر از محتویات این کتاب بدون اجازه کتبی ناشر ممنوع است. متخلفان به موجب قانون حمایت حقوق مؤلفان، مصنفان و هنرمندان تحت پیگرد قانونی قرار می‌گیرند.

◀ عنوان کتاب: آموزش داده کاوی با Rapidminer

◀ مولف: ساسان کرمی زاده

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ ویراستار: مهدیه مخبری

◀ صفحه آرای: نازنین نصیری

◀ طراح جلد: داریوش فرسای

◀ نوبت چاپ: اول

◀ تاریخ نشر: ۱۴۰۲

◀ چاپ و صحافی: صدف

◀ تیراژ: ۱۰۰ جلد

◀ قیمت: ۱۶۸۰۰۰۰ ریال

◀ شابک: ۹۷۸-۶۲۲-۲۱۸-۷۰۶-۴

نشانی واحد فروش: تهران، خیابان انقلاب، خیابان دانشگاه

-تقاطع شهدای ژاندارمری-پلاک ۱۵۸ ساختمان دانشگاه-

طبقه دوم-واحد ۴ تلفن ها: ۶۶۹۶۵۷۴۹-۲۲۰۸۵۱۱۱

فروشگاههای اینترنتی دیباگران تهران :

WWW.MFTBOOK.IR

www.dibagaran-tehran.com

سرشناسه: کرمی زاده، ساسان، ۱۳۶۳-
عنوان و نام پدیدآور: آموزش داده کاوی با Rapidminer
مولف: ساسان کرمی زاده؛
ویراستار: مهدیه مخبری.
مشخصات نشر: تهران: دیباگران تهران: ۱۴۰۲
مشخصات ظاهری: ۱۶۰ ص: مصور، جدول، نمودار.
شابک: ۹۷۸-۶۲۲-۲۱۸-۷۰۶-۴
وضعیت فهرست نویسی: فیپا یادداشت: کتابنامه: ص: ۱۶۰
موضوع: زبیدماینر (منبع الکترونیکی)
موضوع: Rapidminer (electronic resource)
موضوع: داده کاوی-نرم افزار
موضوع: data mining-software
رده بندی کنگره: ۷۶/۹ QA
رده بندی دیویی: ۰۰۶/۳۱۲
شماره کتابشناسی ملی: ۹۲۷۱۵۸۱

نشانی تلگرام: @mftbook نشانی اینستاگرام دیبا dibagaran_publishing

هر کتاب دیباگران، یک فرصت جدید علمی و شغلی.

هر گوشی همراه، یک فروشگاه کتاب دیباگران تهران.

از طریق سایتهای دیباگران، در هر جای ایران به کتابهای ما دسترسی دارید.

فهرست مطالب

مقدمه ناشر ۶

فصل اول

آموزش کار با داده‌ها ۷

- ۱-۱- معرفی نرم‌افزار ریپید ماینر ۸
- ۱-۱-۱- نرم‌افزار ریپید ماینر چیست ۹
- ۲-۱- آشنایی با محیط نرم‌افزار ۱۰
- ۱-۲-۱- مؤلفه‌ها ۱۱
- ۲-۲-۱- عملگرها ۱۱
- ۳-۱- مراحل داده‌کاوی ۴۳
- ۱-۳-۱- پیش‌پردازش داده ۴۵
- ۲-۳-۱- یادگیری ۴۶
- ۳-۳-۱- ارزیابی و تفسیر مدل ۴۷
- ۴-۱- ورود داده‌ها ۴۷
- ۵-۱- فیلترینگ و مرتب کردن داده‌ها ۴۷
- ۶-۱- گروه‌بندی داده‌ها ۵۰
- ۷-۱- ایجاد ویژگی جدید ۵۴
- ۸-۱- آماده‌سازی داده‌ها ۵۶
- ۹-۱- تعریف Role برای ویژگی‌ها ۵۷

فصل دوم

پیش‌پردازنده داده‌ها ۵۹

- ۱-۲- پیش‌بینی ۵۹
- ۲-۲- جای‌گذاری داده‌های مفقود ۶۰
- ۳-۲- نرمال‌سازی داده‌ها ۶۶
- ۴-۲- کاهش ابعاد داده‌ها ۶۸
- ۵-۲- استفاده از ماکروها ۷۳
- ۶-۲- Looping & Branching ۷۷
- ۷-۲- ذخیره داده‌های آماری و نتیجه ۸۶

فصل سوم

۸۷ مدل سازی در رییدمایننز

- ۸۷ ۱-۳- مدل سازی داده
- ۸۹ ۲-۳- اجزاء مدل سازی داده
- ۹۰ ۱-۲-۳- موجودیت
- ۹۰ ۲-۲-۳- ارتباط
- ۹۱ ۳-۲-۳- ویژگی
- ۹۲ ۳-۳- اعتبارسنجی
- ۹۳ ۴-۳- روش های مختلف اعتبارسنجی متقابل
- ۹۳ ۱-۴-۳- روش اعتبارسنجی متقابل (k-fold)
- ۹۴ ۲-۴-۳- اعتبارسنجی ضربدری
- ۹۴ ۳-۴-۳- نمونه گیری چندباره تصادفی
- ۹۵ ۴-۴-۳- یکی-بیرون
- ۹۵ ۵-۴-۳- Holdout روش
- ۹۵ ۶-۴-۳- اعتبارسنجی مدل ها با استفاده از اعتبارسنجی تقسیم
- ۹۷ ۵-۳- محاسبه محتمل ترین حالت ویژگی
- ۱۰۰ ۶-۳- محاسبه شاخص های مرکزی
- ۱۰۵ ۷-۳- امید ریاضی
- ۱۱۸ ۸-۳- تست نتایج مدلسازی و اعتبارسنجی نتایج
- ۱۲۲ ۹-۳- ساخت یک مدل در ریید ماینر

فصل چهارم

۱۲۸ مصورسازی در رییدماینر

- ۱۲۸ ۱-۴- مصورسازی داده
- ۱۳۰ ۲-۴- ویژگی های نمایش تصویری مؤثر
- ۱۳۲ ۳-۴- مصورسازی داده با نمایش اطلاعات عددی
- ۱۳۵ ۴-۴- درک تصویری و مصورسازی داده
- ۱۳۵ ۵-۴- ادراک، شناخت و تجسم در انسان
- ۱۴۰ ۶-۴- مثالی از مصورسازی در ریید ماینر

فصل پنجم

۱۴۴ خوشه بندی در رییدماینر

- ۱۴۴ ۱-۵- منظور از خوشه بندی چیست؟

- ۱۴۷..... ۲-۵- چه تفاوتی بین خوشه‌بندی و دسته‌بندی
- ۱۴۸..... ۳-۵- مدل‌های خوشه‌بندی
- ۱۴۸..... ۱-۳-۵- خوشه‌بندی مبتنی بر افراز
- ۱۴۹..... ۲-۳-۵- خوشه‌بندی تفکیکی
- ۱۵۰..... ۳-۳-۵- خوشه‌بندی سلسله‌مراتبی
- ۱۵۲..... ۴-۳-۵- خوشه‌بندی مبتنی بر تراکم یا چگالی داده‌ها
- ۱۵۲..... ۵-۳-۵- خوشه‌بندی مبتنی بر شبکه‌های شطرنجی گرید
- ۱۵۲..... ۶-۳-۵- خوشه‌بندی مبتنی بر مدل
- ۱۵۳..... ۷-۳-۵- خوشه‌بندی پارتیشن‌بندی
- ۱۵۴..... ۸-۳-۵- خوشه‌بندی مرکز محور
- ۱۵۵..... ۴-۵- ارزیابی مدل‌های خوشه‌بندی
- ۱۵۶..... ۵-۵- برخی از کاربردهای خوشه‌بندی در داده‌کاوی در دنیای واقعی
- ۱۵۶..... ۱-۵-۵- شناسایی اخبار جعلی
- ۱۵۶..... ۲-۵-۵- فیلترکردن ایمیل‌های هرزنامه
- ۱۵۶..... ۳-۵-۵- بازاریابی و فروش
- ۱۵۷..... ۴-۵-۵- شناسایی ترافیک شبکه
- ۱۵۷..... ۵-۵-۵- تجزیه و تحلیل اسناد
- ۱۵۷..... ۶-۵- پیاده‌کردن الگوریتم خوشه‌بندی k-means با استفاده از نرم‌افزار ریپید ماینر
- ۱۵۷..... ۱-۶-۵- تفسیر داده‌ها
- ۱۵۸..... ۲-۶-۵- الگوریتم خوشه‌بندی K-means
- ۱۵۹..... ۳-۶-۵- مراحل پردازشی
- ۱۶۰..... منابع

خط‌مشی انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب‌هایی با کیفیت عالی است که بتواند
خواسته‌های به روز جامعه فرهنگی و علمی کشور را تا حد امکان پوشش دهد.
هر کتاب دیباگران تهران، یک فرصت جدید شغلی و علمی

حمد و سپاس ایزد منان را که با الطاف بی‌کران خود این توفیق را به ما ارزانی داشت تا بتوانیم در راه ارتقای دانش عمومی و فرهنگی این مرز و بوم در زمینه چاپ و نشر کتب علمی و آموزشی گام‌هایی هرچند کوچک برداشته و در انجام رسالتی که بر عهده داریم، مؤثر واقع شویم.

گسترده‌گی علوم و سرعت توسعه روزافزون آن، شرایطی را به وجود آورده که هر روز شاهد تحولات اساسی چشمگیری در سطح جهان هستیم. این گسترش و توسعه، نیاز به منابع مختلف از جمله کتاب را به عنوان قدیمی‌ترین و راحت‌ترین راه دستیابی به اطلاعات و اطلاع‌رسانی، بیش از پیش برجسته نموده است.

در این راستا، واحد انتشارات مؤسسه فرهنگی هنری دیباگران تهران با همکاری اساتید، مؤلفان، مترجمان، متخصصان، پژوهشگران و محققان در زمینه‌های گوناگون و مورد نیاز جامعه تلاش نموده برای رفع کمبودها و نیازهای موجود، منابعی پُر بار، معتبر و با کیفیت مناسب در اختیار علاقمندان قرار دهد.

کتابی که در دست‌دارید تألیف "جناب آقای دکتر ساسان کرمی زاده" است که با تلاش همکاران ما در نشر دیباگران تهران منتشر گشته و شایسته است از یکایک این گرامیان تشکر و قدردانی کنیم.

با نظرات خود مشوق و راهنمای ما باشید

با ارائه نظرات و پیشنهادات و خواسته‌های خود، به ما کمک کنید تا بهتر و دقیق‌تر در جهت رفع نیازهای علمی و آموزشی کشورمان قدم برداریم. برای رساندن پیام‌هایتان به ما از رسانه‌های دیباگران تهران شامل سایتهای فروشگاهی و صفحه اینستاگرام و شماره‌های تماس که در صفحه شناسنامه کتاب آمده استفاده نمایید.

مدیر انتشارات

مؤسسه فرهنگی هنری دیباگران تهران
dibagaran@mftplus.com