



مؤسسه فرهنگی هنری
دییگران تهران

به نام خدا



مؤسسه فرهنگی هنری
دییگران تهران

آموزش داده کاوی با Rapid Miner

مؤلف

دکتر محمد بهار لوئی



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 می گیرند.

◀ عنوان کتاب: آموزش داده کاوی با Rapid Miner

◀ مولف : دکتر محمد بهار لوتی

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ صفحه آرای: فرناوش عبدالمهی

◀ طراح جلد: داریوش فرسای

◀ نوبت چاپ: اول

◀ تاریخ نشر: ۱۳۹۸

◀ چاپ و صحافی: درج عقیق

◀ تیراژ: ۱۰۰ جلد

◀ قیمت: ۳۵۰۰۰۰ ریال

◀ شابک: ۹۷۸-۶۲۲-۲۱۸-۲۱۴-۴

◀ نشانی واحد فروش: تهران، میدان انقلاب،

خ کارگر جنوبی، روبروی پاساژ مهستان،

پلاک ۱۲۵۱

تلفن: ۲۲۰۸۵۱۱۱-۶۶۴۱۰۰۴۶

◀ فروشگاههای اینترنتی دیباگران تهران :

WWW.MFTBOOK.IR

www.dibbook.ir

www.dibagarantehran.com

◀ نشانی تلگرام: @mftbook

◀ نشانی اینستاگرام دیبا dibagaran_publishing

هر کتاب دیباگران، یک فرصت جدید شغلی.

هر گوشی همراه، یک فروشگاه کتاب دیباگران تهران.

از طریق سایتها و اپ دیباگران، در هر جای ایران به کتابهای ما دسترسی دارید.

فهرست مطالب

مقدمه ۹

فصل اول معرفی داده کاوی

مقدمه	۱۲
۱-۱- مفهوم داده کاوی و کشف دانش	۱۲
۲-۱- مراحل داده کاوی	۱۶
۱-۲-۱- پیش پردازش داده	۱۶
۲-۲-۱- یادگیری	۱۸
۳-۲-۱- تفسیر نتایج	۲۰

فصل دوم آشنایی با نرم افزار

مقدمه	۲۲
۱-۲- نصب نرم افزار	۲۲
۲-۲- آشنایی با محیط نرم افزار	۲۷
۱-۲-۲- گزینه های اصلی	۲۸
۲-۲-۲- نوار ابزار	۲۸
۳-۲-۲- صفحه طراحی	۲۹
۴-۲-۲- عملگرها	۳۰
۵-۲-۲- پارامترها	۳۱

فصل سوم

عملگرها

۳۴	مقدمه
۳۴	۱-۳- وارد کردن داده
۴۳	۲-۳- پیش پردازش داده
۴۳	Replace Missing Value - ۱-۲-۳
۴۷	Select attribute - ۲-۲-۳
۴۹	Filter Example - ۳-۲-۳
۵۲	۴-۲-۳- عملگرهای دیگر

فصل چهارم

خوشه بندی در داده کاوی

۵۴	مقدمه
۵۵	۱-۴- ورود داده ها و پیش پردازش
۵۸	۲-۴- خوشه بندی داده ها
۶۰	۳-۴- تفسیر نتایج
۶۳	۴-۴- عملگرهای خاص
۶۳	Write Excel - ۱-۴-۴
۶۵	Nominal to Numeric - ۲-۴-۴

فصل پنجم
دسته بندی در داده کاوی

۶۸..... مقدمه

۶۹..... ۱-۵- ورود و پیش پردازش داده

۷۲..... ۲-۵- رسم درخت تصمیم

۷۴..... ۳-۵- تفسیر مدل

۷۵..... ۴-۵- حالت‌های خاص

۷۵..... ۱-۴-۵- Set role

۷۶..... کلام آخر

۷۷..... منابع

خط مشی کیفیت انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب‌های است که بتواند
خواسته‌های به روز جامعه فرهنگی و علمی کشور را تا حد امکان پوشش دهد.
هر کتاب دیباگران تهران، یک فرصت جدید شعلی

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 دانشگاهی، علوم پایه و به ویژه علوم کامپیوتر و انفورماتیک گام‌هایی هرچند کوچک برداشته و در انجام رسالتی که بر عهده داریم، مؤثر واقع شویم.

گسترده‌گی علوم و 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 روشن می‌نماید.

در این راستا، واحد انتشارات مؤسسه فرهنگی هنری دیباگران تهران با همکاری جمعی از اساتید، مؤلفان، مترجمان، متخصصان، پژوهشگران، محققان و نیز پرسنل ورزیده و ماهر در زمینه امور نشر درصدد هستند تا با تلاش‌های مستمر خود برای رفع کمبودها و نیازهای موجود، منابعی پُر بار، معتبر و با کیفیت مناسب در اختیار علاقمندان قرار دهند.

کتابی که در دست دارید با همت "دکتر محمد بهار لویی" و تلاش جمعی از همکاران انتشارات میسر گشته که شایسته است از یکایک این گرامیان تشکر و قدردانی کنیم.

کارشناسی و نظارت بر محتوا: زهره قزلباش

در خاتمه ضمن سپاسگزاری از شما دانش‌پژوه گرامی درخواست می‌نمایم با مراجعه به آدرس dibagaran.mft.info (ارتباط با مشتری) فرم نظرسنجی را برای کتابی که در دست دارید تکمیل و ارسال نموده، انتشارات دیباگران تهران را که جلب رضایت و وفاداری مشتریان را هدف خود می‌داند، یاری فرمایید.

امیدواریم همواره بهتر از گذشته خدمات و محصولات خود را تقدیم حضورتان نماییم.

مدیر انتشارات

مؤسسه فرهنگی هنری دیباگران تهران
bookmarket@mft.info



مقدمه مولف

بالا رفتن تعداد داده های موجود در سازمانها و همچنین داده های محیطی، فرآیند تصمیم گیری و همچنین پیش بینی آینده را با مشکلات متعددی روبرو کرده است که از جمله آنها می توان به ابهام و سردرگمی مدیران و تصمیم گیرندگان اشاره کرد. در این شرایط، ابزارها و روشهای کیفی و کمی متعددی معرفی شده است که از جمله آنها می توان به داده کاوی اشاره کرد. داده کاوی یکی از مسیرهای دستیابی به دانش می باشد و قادر است تا از داده های در دسترس دانش کاربردی و قابل استفاده را فراهم سازد.

از آنجائیکه داده کاوی حجم بالایی از داده ها را با استفاده از الگوریتم های پیچیده تجزیه و تحلیل می کند، بنابراین برای انجام آن نیاز به کاربست نرم افزار می باشد و بر همین اساس نیز نرم افزارهای متعددی در چند سال گذشته وارد بازار داده کاوی شده است. از جمله نرم افزارهای داده کاوی که در بین مدیران، تصمیم گیرندگان و دانشجویان محبوبیت بالایی دارد، ریپید ماینر می باشد. نرم افزار ریپید ماینر توسط شرکتی با همین نام طراحی شده است و در چند سال گذشته نسخه های بسیاری از آن معرفی و در دسترس کاربران قرار گرفته است. سهولت استفاده، قابلیت های فراوان و محیط کاربر پسند سه ویژگی اصلی این نرم افزار است که باعث بالا رفتن نرخ استفاده از آن و همچنین صدر نشینی این نرم افزار در چندین دوره شده است.

نکته مهم در خصوص نرم افزار ریپید ماینر و به صورت کلی، داده کاوی، دامنه کاربرد آن در رشته های مختلف همچون کامپیوتر، فناوری اطلاعات، مدیریت، اقتصاد، پزشکی و جرم شناسی است. این امر باعث افزایش به کارگیری و همچنین بالا رفتن تعداد کاربران آن شده است. با توجه به اهمیت و علاقه کاربران و همچنین نیاز جدی بسیاری از دانشجویان و تحلیل گران به یادگیری این نرم افزار، آموزشهای متعددی به شکل کتاب، آموزشهای چند رسانه ای و یا کلاسهای آموزشی فراهم شده است. در ادامه این موارد، کتاب حاضر نیز قصد دارد تا آموزشی ساده و قابل درک برای تمامی علاقه مندان فراهم آورد.

هدف اصلی این کتاب افرادی است که حداقل دانش را در زمینه علوم تحلیل داده دارند. به بیان دیگر، این کتاب درصدد آن است تا موضوع داده کاوی به کمک نرم افزار را به صورت کاملاً ساده، به گونه ای که برای کلیه افراد قابل درک باشد، آموزش دهد. بر همین اساس ممکن است که برخی موارد به صورت کاملاً حرفه ای و تخصصی بیان نشود و یا از بیان جزئیات بسیار دقیق صرف نظر شده باشد، اما بدیهی است که بیان ساده کلیه مطالب می تواند یادگیری این علم را با کمک نرم افزار ریپید ماینر برای مبتدیان فراهم سازد.

امید است تا مطالعه این کتاب بتواند دانش داده کاوی و همچنین چگونگی استفاده از نرم افزار ریپید ماینر را به شکلی کاملاً ساده و قابل درک برای تمام علاقه مندان به این علم فراهم سازد. بدیهی است در طول مطالعه این کتاب و در جریان استفاده از نرم افزار، ممکن است مشکلات و خطاهای فراوانی به وجود بیاید. برای رفع این مشکلات می توانید از کتابهای تخصصی تر و جامع تر استفاده کرده و یا سوال خود را از طریق ارسال رایانامه به آدرس mohammadbaharlouei@gmail.com مطرح سازید.