

داده کاوی کاربردی

(ویراست دوم)

نویسندگان:

دکتر محمد صنیعی آباده

(عضو هیأت علمی دانشگاه تربیت مدرس)

مهندس سینا محمودی

مهندس محدثه طاهرپور

نیاز دانش

پیشگفتار

امروزه شاهد تولید حجم عظیمی از داده‌ها در دنیا هستیم. این حجم عظیم داده‌ها در سازمان‌ها، مراکز پزشکی و بیمارستانی، بانکی، بیمه‌ای، بورس، مخابراتی، انرژی، حمل و نقل، بازرگانی، نظامی، امنیتی، سیاسی و غیره تولید می‌شوند. یکی از علوم بسیار استراتژیک که امروزه در دنیا از اهمیت ویژه‌ای برخوردار است و بی‌شک می‌توان نفوذ و پیشرفت روزافزون آن را در بسیاری از کاربردهای علمی و صنعتی مشاهده نمود، علم داده‌کاوی و کشف دانش است. هدف این علم کاوش داده‌ها و استخراج دانش نهفته در آنهاست. دانش تولید شده را می‌توان در بدنه سیستم‌های پشتیبان تصمیم تعبیه نمود و از آن برای تصمیم‌گیری‌های کلیدی استفاده کرد.

حوزه‌های کاربردی علم داده‌کاوی به اندازه وسعت معنای لغوی داده، وسیع بوده و از تنوع بسیار بالایی برخوردار است. به عنوان نمونه می‌توان به حوزه‌های مالی (بانک، بیمه و بورس)، حوزه‌های پزشکی، آزمایشگاهی و بهداشتی، حوزه‌های سیاسی، اطلاعاتی، امنیتی و نظامی، حوزه‌های ارتباطاتی و مخابراتی، حوزه‌های بازاریابی و فروش اشاره نمود. برخی از کاربردهای داده‌کاوی در هر کدام از حوزه‌های مزبور عبارت است از:

- در حوزه‌های مالی:
 - تشخیص نظم حاکم بر تقلب‌های رخ داده در حوزه‌های بانکی، بیمه و بورس.
 - تعیین اعتبار مشتریان آینده جهت اعطای تسهیلات مالی با توجه به رکورد مشتریان گذشته.
 - پیش‌بینی تغییرات آینده در شاخص کلی و همچنین سهام شرکت‌های گوناگون در بورس اوراق بهادار.
 - پیش‌بینی خسارت‌های آتی در پرونده‌های بیمه‌ای.

- در حوزه‌های پزشکی، آزمایشگاهی و بهداشتی:
 - کشف نظم حاکم بر داده‌های بیماری‌های مختلف و تولید سیستم‌های تصمیم‌یار پزشکی قابل استفاده در بیمارستان‌ها، کلینیک‌ها و مراکز آموزش پزشکی.
 - توصیف داده‌های پزشکی در راستای تولید دارو در مراکز آزمایشگاهی.
 - تحلیل تخلف‌های پزشکی رخ داده در مراکز مختلف و تولید سیستم تشخیص‌دهنده تخلف.

- در حوزه‌های سیاسی، اطلاعاتی، امنیتی و نظامی:
 - تحلیل داده‌های مربوط به کنگره‌ها، همایش‌ها و جلسات سیاسی و پیش‌بینی رفتار و رویکرد سیاسی کشورهای مختلف در شرایط گوناگون.
 - کاوش داده‌های جنگی و اتخاذ تصمیمات استراتژیک نظامی.
 - تحلیل داده‌های پروژه‌های مختلف و تولید نظم حاکم بر استراتژی سازمان‌های اطلاعاتی کشورها در رویدادهای گوناگون.

- در حوزه‌های ارتباطاتی و مخابراتی:
 - تحلیل داده‌های تماس‌های تلفنی و کشف نظم‌های حاکم بر آن‌ها.
 - کاوش داده‌های تراکنش‌های تقلبی و استفاده از آن برای کشف سوء استفاده‌های آتی.
 - تحلیل داده‌های مربوط به مشکلات و خرابی‌های شبکه‌های مخابراتی و ارتباطاتی.

- در حوزه‌های بازاریابی و فروش:
 - تحلیل داده‌های خرید مشتریان در فروشگاه‌ها و مراکز تجاری و کشف دانش نهفته در الگوهای رفتاری مشتریان.
 - کاوش داده‌های مربوط به توزیع و پخش کالا در حوزه‌های گوناگون نظیر داده‌های مواد غذایی، پوشاک، بهداشتی و غیره.

این کتاب در دو بخش به طرح مباحث نظری و عملی داده‌کاوی می‌پردازد. در بخش اول که از چهار فصل تشکیل شده است، خواننده کتاب با مفاهیم نظری داده‌کاوی آشنا می‌شود. فصل اول به معرفی داده‌کاوی و کشف دانش پرداخته و تعاریف مقدماتی، اصول، الگوریتم‌ها و کاربردهای متنوع آن را ارائه می‌نماید. فصل دوم به مرحله نخست پردازشی در داده‌کاوی یعنی آماده‌سازی و پویس داده می‌پردازد. در این فصل خواننده با کلیه مراحل که داده را برای ارائه به مرحله بعدی یعنی یادگیری مدل آماده می‌نماید، آشنا می‌شود. در فصل سوم روش‌ها و رویکردهای مختلف یادگیری مدل در داده‌کاوی تشریح شده و الگوریتم‌های گوناگون دسته‌بندی، خوشه‌بندی و کاوش قوانین انجمنی ارائه می‌شوند. فصل چهارم به بررسی روش‌های متنوع ارزیابی مدل پرداخته، و تحلیلی جامع در این زمینه ارائه می‌نماید. بخش دوم کتاب مربوط به فصولی است که از نگاهی عملی به آزمایش و ارزیابی روش‌ها، الگوریتم‌ها و رویکردهای مراحل مختلف داده‌کاوی که در بخش نخست کتاب بحث شده‌اند، می‌پردازد. فصل پنجم به تشریح ابزارها و نرم‌افزارهای مهم و مشهور داده‌کاوی پرداخته و از میان آنها نرم‌افزار

داده کاوی RapidMiner را به عنوان ابزار برتر داده کاوی در دنیا معرفی کرده و به شکلی عمیق تر بررسی می نماید. در فصل ششم مراحل سه گانه کاوش داده که در فصول دوم، سوم و چهارم تشریح شدند، با ذکر چندین مثال عملی به صورتی کاربردی ارائه می شود. نکته مهم و جالب توجهی که می بایست به آن اشاره نمود آن است که در ارائه فصل ششم توجه شده است که ترتیب مراحل و مطالب این فصل سنخیت کاملی با مطالب فصول دوم، سوم و چهارم کتاب داشته باشد به گونه ای که امکان مراجعه مستقیم خواننده محترم کتاب به مطالب عملی در فصل ششم پس از مطالعه مطالب نظری مرتبط در فصول بخش اول میسر باشد. در دو فصل آخر یعنی فصول هفتم و هشتم، لیست کاملی از عملگرهای دارای کاربرد فراوان نرم افزار RapidMiner به همراه توضیحات آن ها ارائه شده است. استفاده از مطالب مربوط به دو فصل مزبور برای خوانندگانی توصیه می شود که آشنایی بالایی با نرم افزار RapidMiner داشته و تنها قصد بررسی و درک عملکرد عملگر مورد نظر خود را دارند. به همین دلیل مطالب دو فصل انتهایی کتاب را می توان به نوعی مرجع عملگرهای نرم افزار RapidMiner دانست.

به یقین کتاب حاضر خالی از اشکال و خطا نیست. پس مزید امتنان است تلاش خوانندگانی که ما را از این اشکالات و خطاها از طریق ناشر مطلع می نمایند. در انتها امیدواریم که مطالب این کتاب بتواند بخشی کوچک از نیاز جامعه علمی و عملی کشور به علم داده کاوی را مورد توجه قرار داده و نیز بستر مناسبی را برای استفاده هر چه مؤثرتر از این علم برای خوانندگان محترم فراهم سازد.

مؤلفین

۱۳۹۳

فهرست مطالب

صفحه	عنوان
۱۷	بخش اول: داده‌کاوی نظری
۱۷	مقدمه
۱۹	فصل اول / معرفی داده‌کاوی و کشف دانش
۱۹	۱-۱ مقدمه
۲۰	۲-۱ انگیزه‌های کاوش داده
۲۱	۱-۲-۱ انگیزه‌های تجاری
۲۳	۲-۲-۱ انگیزه‌های علمی
۲۴	۳-۱ چالش‌های داده‌کاوی
۲۴	۱-۳-۱ چالش‌های اولیه
۲۶	۲-۳-۱ چالش‌های ثانویه
۲۶	۴-۱ معرفی داده‌کاوی
۲۸	۱-۴-۱ منشأ علمی
۲۸	۲-۴-۱ مراحل داده‌کاوی
۲۹	۱-۲-۴-۱ آماده‌سازی داده
۳۰	۲-۲-۴-۱ یادگیری مدل
۳۰	۳-۲-۴-۱ ارزیابی و تفسیر مدل
۳۱	۵-۱ روش‌های یادگیری مدل در داده‌کاوی
۳۱	۱-۵-۱ روش‌های پیش‌بینی
۳۱	۱-۱-۵-۱ دسته‌بندی
۳۵	۲-۱-۵-۱ رگرسیون
۳۶	۳-۱-۵-۱ تشخیص انحراف
۳۶	۲-۵-۱ روش‌های توصیفی
۳۷	۱-۲-۵-۱ خوشه‌بندی
۳۹	۲-۲-۵-۱ کشف قوانین انجمنی
۴۱	۳-۲-۵-۱ کشف الگوهای ترتیبی
۴۲	۶-۱ خلاصه

۴۳	تمرین
۴۳	مراجع

فصل دوم / آماده‌سازی و پویش داده ۴۵

۴۵	۱-۲ مقدمه
۴۶	۲-۲ انواع داده
۴۶	۱-۲-۲ بعد ویژگی
۴۸	۲-۲-۲ بعد تعداد مقادیر
۴۸	۳-۲-۲ بعد مجموعه داده
۴۹	۱-۳-۲-۲ داده‌های رکوردی
۵۱	۲-۳-۲-۲ داده‌های گرافی
۵۱	۳-۳-۲-۲ داده‌های ترتیبی
۵۲	۳-۲ آماده‌سازی داده
۵۲	۱-۳-۲ استخراج داده
۵۲	۱-۱-۳-۲ جمع‌آوری داده
۵۳	۲-۱-۳-۲ انتخاب داده
۵۳	۲-۳-۲ پیش‌پردازش داده
۵۳	۱-۲-۳-۲ پاک‌سازی داده
۵۶	۲-۲-۳-۲ انتخاب زیرمجموعه ویژگی‌ها
۵۷	۳-۲-۳-۲ فیلترینگ نمونه‌ها
۵۸	۴-۲-۳-۲ نمونه‌برداری
۶۰	۵-۲-۳-۲ تبدیل داده
۶۰	۶-۲-۳-۲ گسسته‌سازی
۶۱	۷-۲-۳-۲ کاهش ابعاد
۶۲	۸-۲-۳-۲ انبوهش داده
۶۳	۹-۲-۳-۲ خلق ویژگی
۶۴	۴-۲ پویش داده
۶۶	۱-۴-۲ آمارهای کلان
۶۸	۲-۴-۲ مصورسازی داده
۶۸	۱-۲-۴-۲ هیستوگرام
۷۰	۲-۲-۴-۲ نمودار جعبه‌ای
۷۱	۳-۲-۴-۲ نمودار پراکندگی
۷۲	۴-۲-۴-۲ نمودار بدنه‌ای
۷۳	۵-۲-۴-۲ نمودارهای ماتریسی
۷۵	۶-۲-۴-۲ نمودار مختصات‌های موازی

۷۶ ۷-۲-۴-۲ نمودار ستاره‌ای
۷۷ ۸-۲-۴-۲ صورتک‌های چرنف
۷۷ ۳-۴-۲ پردازش تحلیلی برخط
۸۰ ۵-۲ خلاصه
۸۱ تمرین
۸۳ مراجع

۸۵ فصل سوم / یادگیری مدل

۸۵ ۱-۳ مقدمه
۸۵ ۲-۳ دسته‌بندی
۸۵ ۱-۲-۳ مفاهیم اصلی
۸۸ ۲-۲-۳ درخت تصمیم
۱۰۵ ۳-۲-۳ دسته‌بندهای مبتنی بر قانون
۱۱۵ ۴-۲-۳ استدلال مبتنی بر حافظه
۱۱۸ ۵-۲-۳ الگوریتم مبتنی بر نظریه بیز
۱۲۲ ۶-۲-۳ شبکه‌های عصبی
۱۲۳ ۷-۲-۳ ماشین بردار پشتیبان
۱۲۷ ۸-۲-۳ روش‌های تجمیعی
۱۳۰ ۹-۲-۳ مسائل چند دسته‌ای
۱۳۱ ۱۰-۲-۳ یادگیری دسته‌های نامتعادل
۱۳۲ ۳-۳ خوشه‌بندی
۱۳۳ ۱-۳-۳ مفاهیم اصلی
۱۳۸ ۲-۳-۳ خوشه‌بندی افرازی
۱۳۹ ۱-۲-۳-۳ الگوریتم خوشه‌بندی K-Means
۱۴۶ ۲-۲-۳-۳ الگوریتم خوشه‌بندی K-Medoids
۱۴۶ ۳-۲-۳-۳ الگوریتم خوشه‌بندی Bisecting K-Means
۱۴۷ ۴-۲-۳-۳ الگوریتم خوشه‌بندی Fuzzy C-Means
۱۴۸ ۳-۳-۳ خوشه‌بندی سلسله‌مراتبی
۱۴۹ ۱-۳-۳-۳ روش خوشه‌بندی تجمعی
۱۵۳ ۲-۳-۳-۳ روش خوشه‌بندی تقسیمی
۱۵۴ ۴-۳-۳ خوشه‌بندی مبتنی بر چگالی
۱۵۸ ۴-۳ کشف قوانین انجمنی
۱۵۸ ۱-۴-۳ مفاهیم اصلی
۱۶۳ ۲-۴-۳ تولید مجموعه اشیاء مکرر
۱۶۳ ۱-۲-۴-۳ الگوریتم Apriori
۱۶۸ ۲-۲-۴-۳ الگوریتم FP-Growth

۱۷۰ ۳-۴-۳ تولید قانون
۱۷۲ ۵-۳ خلاصه
۱۷۲ تمرین
۱۸۰ مراجع

فصل چهارم / ارزیابی و تفسیر مدل ۱۸۱

۱۸۱ ۱-۴ مقدمه
۱۸۱ ۲-۴ ارزیابی الگوریتم‌های دسته‌بندی
۱۸۱ ۱-۲-۴ معیارهای ارزیابی الگوریتم‌های دسته‌بندی
۱۹۰ ۲-۲-۴ روش‌های ارزیابی الگوریتم‌های دسته‌بندی
۱۹۱ Holdout روش ۱-۲-۲-۴
۱۹۱ Random Subsampling روش ۲-۲-۲-۴
۱۹۲ Cross-Validation روش ۳-۲-۲-۴
۱۹۳ Bootstrap روش ۴-۲-۲-۴
۱۹۳ ۳-۴ ارزیابی الگوریتم‌های خوشه‌بندی
۱۹۳ ۱-۳-۴ معیارهای ارزیابی بدون ناظر
۱۹۹ ۲-۳-۴ معیارهای ارزیابی باناظر
۲۰۱ ۴-۴ ارزیابی الگوریتم‌های کشف قوانین انجمنی
۲۰۸ ۵-۴ تفسیر مدل
۲۰۸ ۶-۴ خلاصه
۲۰۹ تمرین
۲۱۱ مراجع

بخش دوم: داده‌کاوی عملی ۲۱۲

فصل پنجم / معرفی ابزار برتر داده‌کاوی RapidMiner ۲۱۵

۲۱۵ ۱-۵ مقدمه
۲۱۵ ۲-۵ انواع ابزارهای داده‌کاوی
۲۱۶ ۱-۲-۵ ابزارهای با مجوز تجاری
۲۱۹ ۲-۲-۵ ابزارهای متن باز
۲۲۱ ۳-۵ مقایسه ابزارهای مختلف
۲۲۵ ۴-۵ ابزار داده‌کاوی RapidMiner
۲۲۵ ۱-۴-۵ طراحی پردازش تحلیلی با RapidMiner
۲۲۶ ۱-۱-۴-۵ عملکرد و انعطاف پذیری
۲۲۸ ۲-۱-۴-۵ مقیاس پذیری
۲۲۹ ۲-۴-۵ فرمت داده ورودی

۲۳۰	۵-۵ آشنایی با محیط RapidMiner
۲۳۰	۱-۵-۵ نصب نرم‌افزار و تعیین مخزن ابتدایی
۲۳۱	۱-۱-۵-۵ دورنماها و دیده‌ها
۲۳۷	۲-۱-۵-۵ طراحی یک دورنما
۲۳۸	۳-۱-۵-۵ عملگرها و دید مخازن
۲۴۱	۴-۱-۵-۵ دید پردازش‌ها
۲۴۲	۵-۱-۵-۵ عملگرها و پردازش‌ها
۲۵۲	۶-۱-۵-۵ انتخاب‌های دیگر دید پردازش
۲۵۳	۷-۱-۵-۵ دید پارامترها
۲۵۵	۸-۱-۵-۵ دید راهنمایی و نظرها
۲۵۵	۹-۱-۵-۵ دید نظرات
۲۵۶	۱۰-۱-۵-۵ دید مرور کلی
۲۵۷	۱۱-۱-۵-۵ دید مشکلات و رونوشت
۲۶۰	۲-۵-۵ اجرای پردازش‌های تحلیلی با نرم‌افزار RapidMiner
۲۶۰	۱-۲-۵-۵ ایجاد یک فرآیند جدید
۲۶۳	۲-۲-۵-۵ اولین پردازش تحلیلی
۲۷۴	۳-۲-۵-۵ اجرای پردازش‌ها
۲۷۶	۱-۳-۲-۵-۵ نگاهی به نتایج حاصله
۲۷۷	۲-۳-۲-۵-۵ نقاط شکست
۲۷۹	۳-۵-۵ نمایش داده‌ها و نتایج
۲۸۰	۱-۳-۵-۵ System Monitor
۲۸۱	۲-۳-۵-۵ نمایش نتایج
۲۸۲	۱-۲-۳-۵-۵ روش‌های نمایش نتایج
۲۸۴	۳-۳-۵-۵ کپی داده‌ها و دیده‌ها
۲۸۵	۴-۳-۵-۵ قالب‌های نمایش
۲۸۵	۱-۴-۳-۵-۵ متن
۲۸۶	۲-۴-۳-۵-۵ جدول‌ها
۲۹۷	۴-۵-۵ مدیریت داده: مخزن
۲۹۷	۱-۴-۵-۵ مخزن نرم‌افزار RapidMiner
۲۹۹	۲-۴-۵-۵ ایجاد یک مخزن جدید
۲۹۹	۳-۴-۵-۵ استفاده از مخزن
۳۰۰	۱-۳-۴-۵-۵ پردازش و توصیفات مخزن وابسته به آن پردازش
۳۰۱	۴-۴-۵-۵ وارد کردن انواع عناصر داده‌ای به داخل مخزن
۳۰۱	۱-۴-۴-۵-۵ وارد کردن مجموعه داده با کمک ویزاردها

۳۰۲ذخیره داده در مخزن با استفاده از عملگر Store
۳۰۳وارد کردن فرمت‌های دیگر با استفاده از عملگرها
۳۰۴ذخیره عناصر داده‌ای با استفاده از دید نتیجه یا دید پردازش
۳۰۵مدیریت و دسترسی به مخزن
۳۰۶پردازش زمینه
۳۰۷داده و فراداده
۳۰۹انتشار فراداده از مخزن و در طول پردازش
۳۱۱۶-۵ خلاصه
۳۱۱تمرین
۳۱۱مراجع

فصل ششم / داده‌کاوی با RapidMiner ۳۱۳

۳۱۳۱-۶ مقدمه
۳۱۵۲-۶ پیش‌پردازش داده در RapidMiner
۳۱۵۱-۲-۶ انواع داده در RapidMiner
۳۱۶۲-۲-۶ آماده‌سازی داده
۳۱۷۱-۲-۲-۶ بارگذاری داده
۳۱۷۱-۱-۲-۲-۶ داده استخراج شده
۳۲۳۲-۱-۲-۲-۶ داده مجازی
۳۲۵۲-۲-۲-۶ پیش‌پردازش داده
۳۲۵۱-۲-۲-۲-۶ پاک‌سازی داده
۳۳۰۲-۲-۲-۲-۶ انتخاب زیرمجموعه ویژگی
۳۳۳۳-۲-۲-۲-۶ فیلترینگ نمونه‌ها
۳۳۵۴-۲-۲-۲-۶ نمونه‌برداری
۳۳۷۵-۲-۲-۲-۶ تبدیل داده
۳۴۱۶-۲-۲-۲-۶ گسسته‌سازی
۳۴۵۷-۲-۲-۲-۶ کاهش ابعاد
۳۴۷۸-۲-۲-۲-۶ انبوهش داده
۳۵۱۹-۲-۲-۲-۶ خلق ویژگی
۳۵۴۳-۲-۶ مصورسازی داده
۳۶۰۳-۶ یادگیری مدل در RapidMiner
۳۶۰۱-۳-۶ دسته‌بندی
۳۶۰۱-۱-۳-۶ الگوریتم C4.5
۳۶۰۱-۱-۳-۶ آماده‌سازی یادگیری درخت تصمیم
۳۶۵۲-۱-۳-۶ تنظیم پارامترهای درخت تصمیم C4.5

۳۷۰ ۳-۱-۳-۶ داده کاوی پنجره‌ای
۳۷۱ ۲-۱-۳-۶ روش‌های تجمیعی
۳۷۲ ۳-۱-۳-۶ مسائل چند دسته‌ای
۳۷۵ ۴-۱-۳-۶ معیارهای ارزیابی یادگیری دسته‌ها
۳۷۶ ۵-۱-۳-۶ یادگیری دسته‌های نامتعادل
۳۸۰ ۲-۳-۶ خوشه‌بندی
۳۸۰ ۱-۲-۳-۶ الگوریتم K-Means
۳۸۳ ۳-۳-۶ کشف قوانین انجمنی
۳۸۳ ۱-۳-۳-۶ الگوریتم FP-Growth
۳۸۴ ۲-۳-۳-۶ تولید قانون
۳۸۶ ۴-۶ ارزیابی و تفسیر مدل در RapidMiner
۳۸۶ ۱-۴-۶ ارزیابی الگوریتم‌های دسته‌بندی
۳۸۸ ۲-۴-۶ ارزیابی الگوریتم‌های خوشه‌بندی
۳۹۰ ۳-۴-۶ ارزیابی الگوریتم‌های کشف قوانین انجمنی
۳۹۰ ۵-۶ خلاصه
۳۹۱ تمرین
۳۹۲ مراجع

۳۹۳ فصل هفتم / عملگرهای مقدماتی نرم‌افزار RapidMiner

۳۹۳ ۱-۷ مقدمه
۳۹۴ ۲-۷ عملگرهای مقدماتی
۳۹۴ ۱-۲-۷ عملگرهای کنترل پردازش
۳۹۴ ۱-۱-۲-۷ پارامترها
۴۰۱ ۲-۱-۲-۷ چرخه
۴۱۰ ۳-۱-۲-۷ انشعاب
۴۱۳ ۴-۱-۲-۷ عملیات مجموعه‌ای روی ورودی‌های جداگانه
۴۱۶ ۵-۱-۲-۷ سایر عملگرهای کنترل پردازش
۴۱۸ ۲-۲-۷ عملگرهای ابزاری
۴۱۸ ۱-۲-۲-۷ عملگرهای خلق داده
۴۲۰ ۲-۲-۲-۷ سایر عملگرها
۴۲۱ ۳-۲-۷ عملگرهای مخزنی
۴۲۱ Retrieve ۱-۳-۲-۷
۴۲۳ Store ۲-۳-۲-۷
۴۲۳ ۴-۲-۷ عملگرهای بارگذاری کننده انواع داده
۴۲۳ ۱-۴-۲-۷ عملگرهای خواندن فایل‌های داده‌ای

۴۴۵.....	Read URL ۱-۱-۴-۲-۷
۴۴۶.....	عملگرهای خواندن مدل ۲-۴-۲-۷
۴۴۷.....	عملگرهای خواندن ویژگی ۳-۴-۲-۷
۴۴۹.....	عملگرهای خواندن نتیجه ۴-۴-۲-۷
۴۴۹.....	سایر عملگرهای خواندن ۵-۴-۲-۷
۴۵۱.....	عملگرهای ذخیره کننده انواع داده ۵-۲-۷
۴۵۱.....	عملگرهای ذخیره فایل‌های داده‌ای ۱-۵-۲-۷
۴۶۲.....	عملگرهای ذخیره کردن مدل ۲-۵-۲-۷
۴۶۳.....	عملگرهای ذخیره کردن ویژگی ۳-۵-۲-۷
۴۶۵.....	عملگرهای ذخیره کردن نتیجه ۴-۵-۲-۷
۴۶۶.....	سایر عملگرهای ذخیره‌سازی ۵-۵-۲-۷
۴۶۹.....	۳-۷ خلاصه
۴۶۹.....	تمرین
۴۶۹.....	مراجع

فصل هشتم / عملگرهای کاوش داده در نرم‌افزار RapidMiner ۴۷۱.....

۴۷۱.....	۱-۸ مقدمه
۴۷۱.....	۲-۸ آماده‌سازی داده
۴۷۲.....	۱-۲-۸ عملگرهای تغییر نام و نقش داده
۴۷۲.....	Set Role ۱-۱-۲-۸
۴۷۳.....	Exchange Role ۲-۱-۲-۸
۴۷۳.....	Rename ۳-۱-۲-۸
۴۷۳.....	Rename by Replacing ۴-۱-۲-۸
۴۷۴.....	Rename by Generic Names ۵-۱-۲-۸
۴۷۴.....	Rename by Constructions ۶-۱-۲-۸
۴۷۴.....	Rename by Example Values ۷-۱-۲-۸
۴۷۵.....	۲-۲-۸ عملگرهای تغییر نوع داده
۴۷۵.....	۱-۲-۲-۸ عملگرهای گسسته‌سازی
۴۷۸.....	۲-۲-۲-۸ عملگرهای تغییر نوع
۴۷۹.....	۳-۲-۸ عملگرهای تبدیل و کاهش مجموعه ویژگی‌ها
۴۸۰.....	۱-۳-۲-۸ Generation بخش
۴۸۴.....	۲-۳-۲-۸ Transformation بخش
۴۸۶.....	۳-۳-۲-۸ Selection بخش
۴۹۲.....	۴-۲-۸ عملگرهای دست‌کاری مقادیر
۴۹۲.....	۱-۴-۲-۸ Normalize

۴۹۲De-Normalize ۲-۴-۲-۸
۴۹۲ Map ۳-۴-۲-۸
۴۹۳Set data ۴-۴-۲-۸
۴۹۳ عملگرهای پاک‌سازی داده ۵-۲-۸
۴۹۳ عملگرهای شناسایی داده‌های پرت ۱-۵-۲-۸
۴۹۴ عملگرهای جایگزینی مقادیر از دست رفته ۲-۵-۲-۸
۴۹۵ عملگرهای فیلترینگ ۶-۲-۸
۴۹۶ عملگرهای نمونه‌برداری ۱-۶-۲-۸
۴۹۸ سایر عملگرهای فیلترینگ ۲-۶-۲-۸
۴۹۹ عملگرهای مرتب‌سازی داده‌ها ۷-۲-۸
۵۰۰ Sort ۱-۷-۲-۸
۵۰۰ Sort by Pareto Rank ۲-۷-۲-۸
۵۰۰Shuffle ۳-۷-۲-۸
۵۰۰ عملگرهای انبوهش داده ۸-۲-۸
۵۰۱Aggregate ۱-۸-۲-۸
۵۰۲ عملگرهای عملیات مجموعه‌ای ۹-۲-۸
۵۰۲ Append ۱-۹-۲-۸
۵۰۲Join ۲-۹-۲-۸
۵۰۲ Set Minus ۳-۹-۲-۸
۵۰۳ Intersect ۴-۹-۲-۸
۵۰۳ Union ۵-۹-۲-۸
۵۰۳ Superset ۶-۹-۲-۸
۵۰۴ Cartesian Product ۷-۹-۲-۸
۵۰۴ عملگرهای مدل‌سازی در نرم‌افزار RapidMiner ۳-۸
۵۰۵ دسته‌بندی و رگرسیون ۱-۳-۸
۵۰۵ استنتاج‌های درختی ۱-۱-۳-۸
۵۰۸ استنتاج مبتنی بر قوانین ۲-۱-۳-۸
۵۱۰ روش‌های مدل‌سازی Lazy ۳-۱-۳-۸
۵۱۱ مدل‌سازی Bayesian ۴-۱-۳-۸
۵۱۳ آموزش دسته‌بندهای شبکه‌های عصبی مصنوعی ۵-۱-۳-۸
۵۱۵ مدل‌سازی ماشین بردار پشتیبان ۶-۱-۳-۸
۵۱۷ مدل‌سازی Meta ۷-۱-۳-۸
۵۱۹ عملگرهای رگرسیون ۸-۱-۳-۸
۵۲۱ وزن‌دهی ویژگی‌ها ۲-۳-۸
۵۲۱ Weight by Information Gain ۱-۲-۳-۸

۵۲۲Weight by Rule	۲-۲-۳-۸
۵۲۲ Weight by Value Average	۳-۲-۳-۸
۵۲۲ Weight by Deviation	۴-۲-۳-۸
۵۲۳ Weight by Chi Squared Statistic	۵-۲-۳-۸
۵۲۳ Weight by Tree Importance	۶-۲-۳-۸
۵۲۳Weight by User Specification	۷-۲-۳-۸
۵۲۴ Data to Weight	۸-۲-۳-۸
۵۲۴ Weight to data	۹-۲-۳-۸
۵۲۴ خوشه‌بندی و قطعه‌بندی	۳-۳-۳-۸
۵۲۴ K-Means	۱-۳-۳-۸
۵۲۵ K-Means (Kernel)	۲-۳-۳-۸
۵۲۵ K-Medoids	۳-۳-۳-۸
۵۲۵ کاوش مجموعه اشیاء مکرر	۴-۳-۳-۸
۵۲۵ FP-Growth	۱-۴-۳-۸
۵۲۶ Create Association rules	۲-۴-۳-۸
۵۲۶ عملگرهای اعمال مدل روی داده	۵-۳-۳-۸
۵۲۶ Apply Model	۱-۵-۳-۸
۵۲۷ RapidMiner عملگرهای ارزیابی و تفسیر مدل در نرم‌افزار	۴-۸
۵۲۷ الگوریتم‌های ارزیابی دسته‌بندی	۱-۴-۸
۵۲۷ Split Validation	۱-۱-۴-۸
۵۲۸ X-validation	۲-۱-۴-۸
۵۲۸ Bootstrapping validation	۳-۱-۴-۸
۵۲۹ عملگرهای ارزیابی الگوریتم‌های خوشه‌بندی	۲-۴-۸
۵۲۹ Map Clustering on Label	۱-۲-۴-۸
۵۲۹ cluster distance performance	۲-۲-۴-۸
۵۳۰ سایر عملگرهای ارزیابی و تفسیر مدل	۳-۴-۸
۵۳۰ Performance	۱-۳-۴-۸
۵۳۰ Performance(User-Based)	۲-۳-۴-۸
۵۳۱ Performance classification	۳-۳-۴-۸
۵۳۱ خلاصه	۵-۸
۵۳۲ تمرین	
۵۳۲ مراجع	
۵۳۳ واژه‌نامه	

بخش اول

داده کاوی نظری

مقدمه

این بخش از چهار فصل تشکیل شده است و در آن خواننده کتاب با مفاهیم نظری داده کاوی آشنا می شود. فصل اول به معرفی داده کاوی و کشف دانش پرداخته و تعاریف مقدماتی، اصول، الگوریتم ها و کاربردهای متنوع آن را ارائه می نماید. فصل دوم به مرحله نخست پردازشی در داده کاوی یعنی آماده سازی و پویس داده می پردازد. در این فصل خواننده با کلیه مراحل که داده را برای ارائه به مرحله بعدی یعنی یادگیری مدل آماده می نماید، آشنا می شود. این مراحل عبارتند از استخراج و پیش پردازش داده. مراحل متنوع پیش پردازش داده کاوی که در این فصل مورد بررسی قرار خواهند گرفت عبارتند از: پاک سازی داده، انتخاب زیرمجموعه ویژگی، فیلترینگ نمونه ها، نمونه برداری، تبدیل داده، گسسته سازی، کاهش ابعاد، انبوهش داده و خلق ویژگی. در فصل سوم روش ها و رویکردهای مختلف یادگیری مدل در داده کاوی تشریح شده و الگوریتم های گوناگون دسته بندی، خوشه بندی و کاوش قوانین انجمنی ارائه خواهند شد. فصل چهارم به بررسی روش های متنوع ارزیابی مدل پرداخته، و تحلیلی جامع در این زمینه ارائه می نماید. لازم به ذکر است که در برخی از مباحث، مطالب و نیز شکل های ارائه شده در این بخش، از مطالب کتاب Introduction to Data Mining نوشته M. Steinbach, P. Tan و V. Kumar در سال ۲۰۰۶ الهام گرفته شده است.