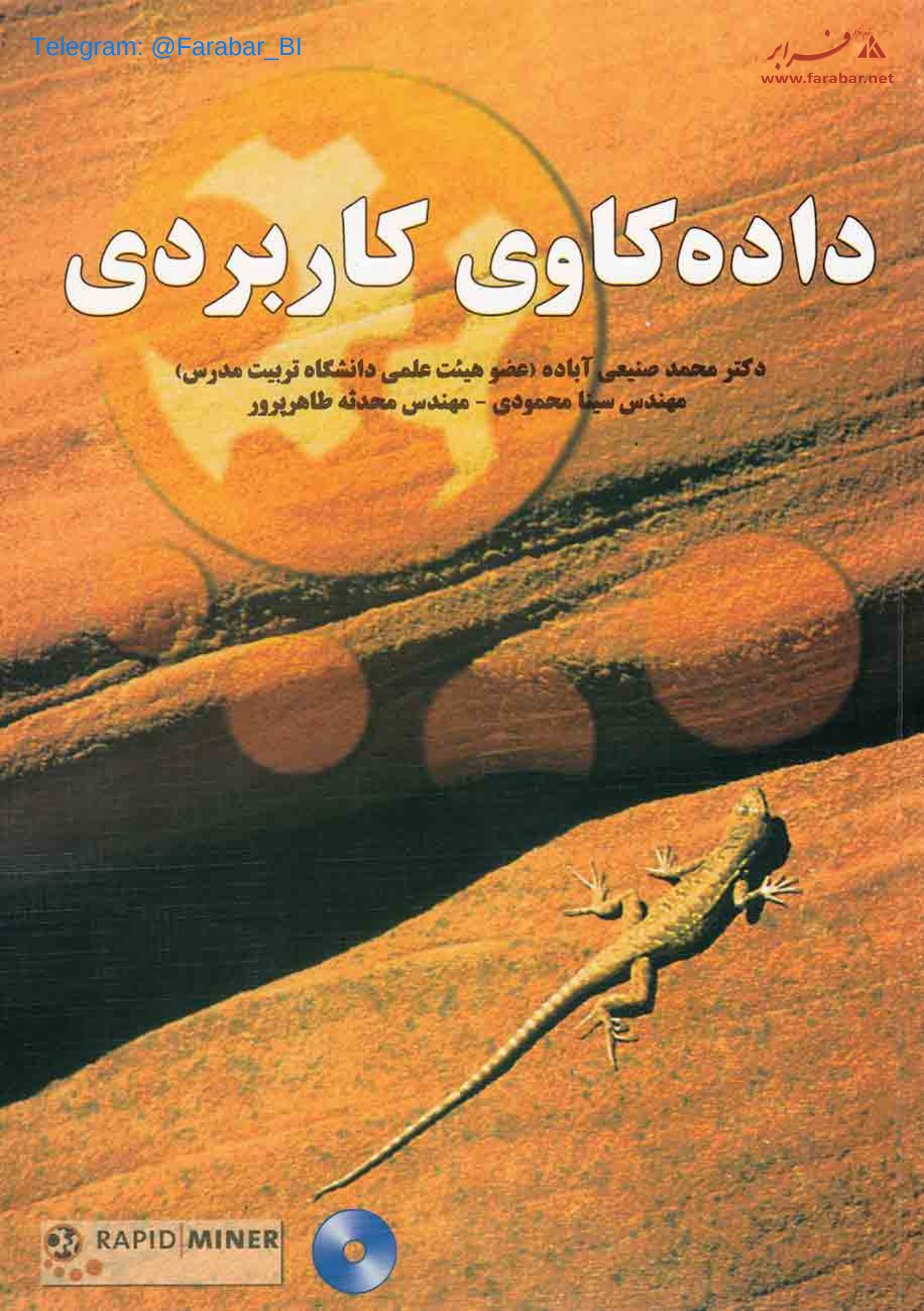
Telegram: @Farabar_BI

فرابر
www.farabar.net

داده‌کاوی کاربردی

دکتر محمد صنیعی آباده (عضو هیئت علمی دانشگاه تربیت مدرس)
مهندس سینا محمودی - مهندس محدثه طاهرپرور



RAPID MINER

Telegram: @Farabar_BI

فرابر
www.farabar.net

سرشناسنامه	:	صنیعی آباده، محمد، ۱۳۵۸ -
عنوان و نام پدیدآور	:	داده‌کاوی کاربردی، نویسندگان محمد صنیعی آباده، سینا محمودی، محدثه طاهرپرور
مشخصات نشر	:	تهران: نیاز دانش، ۱۳۹۱.
مشخصات ظاهری	:	۵۱۲ص.: مصور (رنگی)، جدول، نمودار
شابک	:	978-600-6481-12-8
وضعیت فهرست‌نویسی	:	فیپا
یادداشت	:	واژه‌نامه
موضوع	:	داده‌کاوی
موضوع	:	داده‌کاوی – نرم‌افزار
شناسنامه افزوده	:	محمودی، سینا، ۱۳۶۳ -
شناسنامه افزوده	:	طاهرپرور، محدثه، ۱۳۶۴ -
رده‌بندی کنگره	:	۱۳۹۱ ۹ ص ۲ د QA۷۶/۹
رده‌بندی دیویی	:	۰۰۶/۳۱۲
شماره کتابشناسی ملی	:	۲۷۴۱۵۸۴

نیاز دانش

نام کتاب	:	داده‌کاوی کاربردی
پدیدآورندگان	:	محمد صنیعی آباده – سینا محمودی – محدثه طاهرپرور
مدیر اجرایی – ناظر بر چاپ	:	حمیدرضا امحمد شیرازی – محمد شمس
طراح جلد	:	کیانا آرین
ناشر	:	نیاز دانش
صفحه‌آرا	:	قلم‌نگار – معمّری
لیتوگرافی / چاپ	:	نص – گنج شایگان
نوبت چاپ	:	اول – ۱۳۹۱
شمارگان	:	۱۰۰۰
قیمت	:	۱۶۰۰۰۰ ریال

شابک: ۸-۱۲-۶۴۸۱-۶۰۰-۹۷۸ ISBN: 978-600-6481-12-8

توزیع ۰۹۱۲-۷۰۷۳۹۳۵ - ۶۶۴۰۵۳۷۲ – ۶۶۴۱۲۳۸۵ – ۶۶۴۶۵۶۷۴

مشاوره جهت نشر: ۰۹۱۲ – ۲۱۰۶۷۰۹

کلیه حقوق مادی و معنوی این اثر متعلق به ناشر می‌باشد.

Telegram: @Farabar_BI

فرابر
www.farabar.net

فهرست مطالب

عنوان | صفحه

بخش اول: داده‌کاوی نظری ۱۷

مقدمه ۱۷

فصل اول / معرفی داده‌کاوی و کشف دانش ۱۹

۱-۱ مقدمه ۱۹

۱-۲ انگیزه‌های کاوش داده ۲۰

۱-۲-۱ انگیزه‌های تجاری ۲۱

۱-۲-۲ انگیزه‌های علمی ۲۳

۱-۳ چالش‌های داده‌کاوی ۲۴

۱-۳-۱ چالش‌های اولیه ۲۴

۱-۳-۲ چالش‌های ثانویه ۲۶

۱-۴ معرفی داده‌کاوی ۲۶

۱-۴-۱ منشأ علمی ۲۸

۱-۴-۲ مراحل داده‌کاوی ۲۸

۱-۴-۲-۱ آماده‌سازی داده ۲۹

۱-۴-۲-۲ یادگیری مدل ۳۰

۱-۴-۲-۳ ارزیابی و تفسیر مدل ۳۰

۱-۵ روش‌های یادگیری مدل در داده‌کاوی ۳۱

۱-۵-۱ روش‌های پیش‌بینی ۳۱

۱-۵-۱-۱ دسته‌بندی ۳۱

۱-۵-۱-۲ رگرسیون ۳۵

۱-۵-۱-۳ تشخیص انحراف ۳۶

۱-۵-۲ روش‌های توصیفی ۳۶

۱-۵-۲-۱ خوشه‌بندی ۳۷

۱-۵-۲-۲ کشف قوانین انجمنی ۳۹

Telegram: @Farabar_BI

فرابر
www.farabar.net

۱-۵-۲-۳ کشف الگوهای ترتیبی ۴۱
۱-۶ خلاصه ۴۲
فصل دوم / آماده‌سازی و پویش داده ۴۵
۲-۱ مقدمه ۴۵
۲-۲ انواع داده ۴۶
۲-۲-۱ بعد ویژگی ۴۶
۲-۲-۲ بعد تعداد مقادیر ۴۸
۲-۲-۳ بعد مجموعه داده ۴۸
۲-۲-۳-۱ داده‌های رکوردی ۴۹
۲-۲-۳-۲ داده‌های گرافی ۵۱
۲-۲-۳-۳ داده‌های ترتیبی ۵۱
۲-۳ آماده‌سازی داده ۵۲
۲-۳-۱ استخراج داده ۵۲
۲-۳-۱-۱ جمع‌آوری داده ۵۲
۲-۳-۱-۲ انتخاب داده ۵۳
۲-۳-۲ پیش‌پردازش داده ۵۳
۲-۳-۲-۱ پاک‌سازی داده ۵۳
۲-۳-۲-۲ انتخاب زیرمجموعه ویژگی‌ها ۵۶
۲-۳-۲-۳ فیلترینگ نمونه‌ها ۵۷
۲-۳-۲-۴ نمونه‌برداری ۵۸
۲-۳-۲-۵ تبدیل داده ۶۰
۲-۳-۲-۶ گسسته‌سازی ۶۰
۲-۳-۲-۷ کاهش ابعاد ۶۱
۲-۳-۲-۸ انبوهش داده ۶۲
۲-۳-۲-۹ خلق ویژگی ۶۳
۲-۴ پویش داده ۶۴
۲-۴-۱ آمارهای کلان ۶۶
۲-۴-۲ مصورسازی داده ۶۸
۲-۴-۲-۱ هیستوگرام ۶۸

Telegram: @Farabar_BI

فرابر
www.farabar.net

۲-۴-۲-۲ نمودار جعبه‌ای ۷۰
۲-۴-۲-۳ نمودار پراکندگی ۷۱
۲-۴-۲-۴ نمودار بدنه‌ای ۷۲
۲-۴-۲-۵ نمودارهای ماتریسی ۷۳
۲-۴-۲-۶ نمودار مختصات‌های موازی ۷۵
۲-۴-۲-۷ نمودار ستاره‌ای ۷۶
۲-۴-۲-۸ صورتک‌های چرنف ۷۷
۲-۴-۳ پردازش تحلیلی برخط ۷۷
۲-۵ خلاصه ۸۰

فصل سوم / یادگیری مدل ۸۱

۳-۱ مقدمه ۸۱
۳-۲ دسته‌بندی ۸۱
۳-۲-۱ مفاهیم اصلی ۸۱
۳-۲-۲ درخت تصمیم ۸۴
۳-۲-۳ دسته‌بندهای مبتنی بر قانون ۱۰۲
۳-۲-۴ استدلال مبتنی بر حافظه ۱۱۱
۳-۲-۵ الگوریتم مبتنی بر نظریه بیز ۱۱۴
۳-۲-۶ شبکه‌های عصبی ۱۱۸
۳-۲-۷ ماشین بردار پشتیبان ۱۱۹
۳-۲-۸ روش‌های تجمیعی ۱۲۳
۳-۲-۹ مسائل چند دسته‌ای ۱۲۶
۳-۲-۱۰ یادگیری دسته‌های نامتعادل ۱۲۷
۳-۳ خوشه‌بندی ۱۲۸
۳-۳-۱ مفاهیم اصلی ۱۲۹
۳-۳-۲ خوشه‌بندی افرازی ۱۳۴
۳-۳-۲-۱ الگوریتم خوشه‌بندی K-Means ۱۳۵
۳-۳-۲-۲ الگوریتم خوشه‌بندی K-Medoids ۱۴۲
۳-۳-۲-۳ الگوریتم خوشه‌بندی Bisecting K-Means ۱۴۲
۳-۳-۲-۴ الگوریتم خوشه‌بندی Fuzzy C-Means ۱۴۳
۳-۳-۳ خوشه‌بندی سلسله مراتبی ۱۴۴
۳-۳-۳-۱ روش خوشه‌بندی تجمعی ۱۴۵

Telegram: @Farabar_BI

www.farabar.net

۵-۴ ابزار داده‌کاوی RapidMiner ۲۰۹
۵-۴-۱طراحی پردازش تحلیلی با RapidMiner ۲۰۹
۵-۴-۱-۱عملکرد و انعطاف پذیری ۲۱۰
۵-۴-۱-۲ مقیاس پذیری ۲۱۲
۵-۴-۲ فرمت داده ورودی ۲۱۳
۵-۵ آشنایی با محیط RapidMiner ۲۱۴
۵-۵-۱ نصب نرم‌افزار و تعیین مخزن ابتدایی ۲۱۴
۵-۵-۱-۱ دورنماها و دیدها ۲۱۶
۵-۵-۱-۲ طراحی یک دورنما ۲۲۱
۵-۵-۱-۳ عملگرها و دید مخازن ۲۲۲
۵-۵-۱-۴ دید پردازش‌ها ۲۲۵
۵-۵-۱-۵ عملگرها و پردازش‌ها ۲۲۶
۵-۵-۱-۶ انتخاب‌های دیگر دید پردازش ۲۳۶
۵-۵-۱-۷ دید پارامترها ۲۳۷
۵-۵-۱-۸ دید راهنمایی و نظرها ۲۳۹
۵-۵-۱-۹ دید نظرات ۲۳۹
۵-۵-۱-۱۰ دید مرور کلی ۲۴۰
۵-۵-۱-۱۱ دید مشکلات و رونوشت ۲۴۱
۵-۵-۲ اجرای پردازش‌های تحلیلی با نرم‌افزار RapidMiner ۲۴۴
۵-۵-۲-۱ ایجاد یک فرآیند جدید ۲۴۴
۵-۵-۲-۲ اولین پردازش تحلیلی ۲۴۷
۵-۵-۲-۳ اجرای پردازش‌ها ۲۵۸
۵-۵-۲-۳-۱ نگاهی به نتایج حاصله ۲۶۰
۵-۵-۲-۳-۲ نقاط شکست ۲۶۱
۵-۵-۳ نمایش داده‌ها و نتایج ۲۶۳
۵-۵-۳-۱ System Monitor ۲۶۴
۵-۵-۳-۲ نمایش نتایج ۲۶۵
۵-۵-۳-۲-۱ روش‌های نمایش نتایج ۲۶۶
۵-۵-۳-۳ کپی داده‌ها و دیدها ۲۶۸

Telegram: @Farabar_BI

www.farabar.net

۵-۵-۳-۴ قالب‌های نمایش ۲۶۹
۵-۵-۳-۴-۱ متن ۲۶۹
۵-۵-۳-۴-۲جدول‌ها ۲۷۰
۵-۵-۴ مدیریت داده: مخزن ۲۸۱
۵-۵-۴-۱ مخزن نرم‌افزار RapidMiner ۲۸۱
۵-۵-۴-۲ ایجاد یک مخزن جدید ۲۸۳
۵-۵-۴-۳ استفاده از مخزن ۲۸۳
۵-۵-۴-۳-۱ پردازش و توصیفات مخزن وابسته به آن پردازش ۲۸۴
۵-۵-۴-۴ وارد کردن انواع عناصر داده‌ای به داخل مخزن ۲۸۵
۵-۵-۴-۴-۱ واردکردن مجموعه داده با کمک ویزاردها ۲۸۵
۵-۵-۴-۴-۲ ذخیره داده در مخزن با استفاده از عملگر Store ۲۸۶
۵-۵-۴-۴-۳ وارد کردن فرمت‌های دیگر با استفاده از عملگرها ۲۸۷
۵-۵-۴-۴-۴ ذخیره عناصر داده‌ای با استفاده از دید نتیجه یا دید پردازش ۲۸۸
۵-۵-۴-۵ مدیریت و دسترسی به مخزن ۲۸۹
۵-۵-۴-۶ پردازش زمینه ۲۹۰
۵-۵-۴-۷ داده و فرادداده ۲۹۱
۵-۵-۴-۸ انتشار فرادداده از مخزن و در طول پردازش ۲۹۳
۵-۶ خلاصه ۲۹۵

فصل ششم / داده‌کاوی با RapidMiner ۲۹۷

۶-۱ مقدمه ۲۹۷
۶-۲ پیش‌پردازش داده در RapidMiner ۲۹۹
۶-۲-۱ انواع داده در RapidMiner ۲۹۹
۶-۲-۲ آماده‌سازی داده ۳۰۰
۶-۲-۲-۱ بارگذاری داده ۳۰۱
۶-۲-۲-۱-۱ داده استخراج شده ۳۰۱
۶-۲-۲-۱-۲ داده مجازی ۳۰۷
۶-۲-۲-۲ پیش‌پردازش داده ۳۰۹
۶-۲-۲-۲-۱ پاک‌سازی داده ۳۰۹

Telegram: @Farabar_BI

فرابر
www.farabar.net

۶-۲-۲-۲-۲ انتخاب زیرمجموعه ویژگی ۳۱۴
۶-۲-۲-۲-۳ فیلترینگ نمونه‌ها ۳۱۷
۶-۲-۲-۲-۴ نمونه‌برداری ۳۱۹
۶-۲-۲-۲-۵ تبدیل داده ۳۲۱
۶-۲-۲-۲-۶ گسسته‌سازی ۳۲۵
۶-۲-۲-۲-۷ کاهش ابعاد ۳۲۹
۶-۲-۲-۲-۸ انبوهش داده ۳۳۱
۶-۲-۲-۲-۹ خلق ویژگی ۳۳۵
۶-۲-۳ مصورسازی داده ۳۳۸
۶-۳ یادگیری مدل در RapidMiner ۳۴۴
۶-۳-۱ دسته‌بندی ۳۴۴
۶-۳-۱-۱ الگوریتم C4.5 ..... ۳۴۴
۶-۳-۱-۱-۱ آماده‌سازی یادگیری درخت تصمیم ۳۴۴
۶-۳-۱-۱-۲ تنظیم پارامترهای درخت تصمیم C4.5 ..... ۳۴۹
۶-۳-۱-۱-۳ داده‌کاوی پنجره‌ای ۳۵۴
۶-۳-۱-۲ روش‌های تجمیعی ۳۵۵
۶-۳-۱-۳ مسائل چند دسته‌ای ۳۵۶
۶-۳-۱-۴ معیارهای ارزیابی یادگیری دسته‌ها ۳۵۹
۶-۳-۱-۵ یادگیری دسته‌های نامتعادل ۳۶۰
۶-۳-۲ خوشه‌بندی ۳۶۴
۶-۳-۲-۱ الگوریتم K-Means ۳۶۴
۶-۳-۳ کشف قوانین انجمنی ۳۶۷
۶-۳-۳-۱ الگوریتم FP-Growth ۳۶۷
۶-۳-۳-۲ تولید قانون ۳۶۸
۶-۴ ارزیابی و تفسیر مدل در RapidMiner ۳۷۰
۶-۴-۱ ارزیابی الگوریتم‌های دسته‌بندی ۳۷۰
۶-۴-۲ ارزیابی الگوریتم‌های خوشه‌بندی ۳۷۲
۶-۴-۳ ارزیابی الگوریتم‌های کشف قوانین انجمنی ۳۷۴
۶-۵ خلاصه ۳۷۴

Telegram: @Farabar_BI

www.farabar.net

فصل هفتم / عملگرهای مقدماتی نرم‌افزار RapidMiner ۳۷۵
۷-۱ مقدمه ۳۷۵
۷-۲ عملگرهای مقدماتی ۳۷۶
۷-۲-۱ عملگرهای کنترل پردازش ۳۷۶
۷-۲-۱-۱ پارامترها ۳۷۶
۷-۲-۱-۲ چرخه ۳۸۳
۷-۲-۱-۳ انشعاب ۳۹۲
۷-۲-۱-۴ عملیات مجموعه‌ای روی ورودی‌های جداگانه ۳۹۵
۷-۲-۱-۵ سایر عملگرهای کنترل پردازش ۳۹۸
۷-۲-۲ عملگرهای ابزاری ۴۰۰
۷-۲-۲-۱ عملگرهای خلق داده ۴۰۰
۷-۲-۲-۲ سایر عملگرها ۴۰۲
۷-۲-۳ عملگرهای مخزنی ۴۰۳
۷-۲-۳-۱ Retrieve ۴۰۳
۷-۲-۳-۲ Store ۴۰۴
۷-۲-۴ عملگرهای بارگذاری کننده انواع داده ۴۰۵
۷-۲-۴-۱ عملگرهای خواندن فایل‌های داده‌ای ۴۰۵
۷-۲-۴-۱-۱ Read URL ۴۲۶
۷-۲-۴-۲ عملگرهای خواندن مدل ۴۲۸
۷-۲-۴-۳ عملگرهای خواندن ویژگی ۴۲۹
۷-۲-۴-۴ عملگرهای خواندن نتیجه ۴۳۰
۷-۲-۴-۵ سایر عملگرهای خواندن ۴۳۱
۷-۲-۵ عملگرهای ذخیره کننده انواع داده ۴۳۳
۷-۲-۵-۱ عملگرهای ذخیره فایل‌های داده‌ای ۴۳۳
۷-۲-۵-۲ عملگرهای ذخیره کردن مدل ۴۴۳
۷-۲-۵-۳ عملگرهای ذخیره کردن ویژگی ۴۴۵
۷-۲-۵-۴ عملگرهای ذخیره کردن نتیجه ۴۴۷
۷-۲-۵-۵ سایر عملگرهای ذخیره‌سازی ۴۴۷
۷-۳ خلاصه ۴۵۱

Telegram: @Farabar_BI

فرابر
www.farabar.net

فصل هشتم / عملگرهای کاوش داده در نرم‌افزار RapidMiner ۴۵۳
۸-۱ مقدمه ۴۵۳
۸-۲ آماده‌سازی داده ۴۵۳
۸-۲-۱ عملگرهای تغییر نام و نقش داده ۴۵۴
۸-۲-۱-۱ Set Role ۴۵۴
۸-۲-۱-۲ Exchange Role ۴۵۵
۸-۲-۱-۳ Rename ۴۵۵
۸-۲-۱-۴ Rename by Replacing ۴۵۶
۸-۲-۱-۵ Rename by Generic Names ۴۵۶
۸-۲-۱-۶ Rename by Constructions ۴۵۶
۸-۲-۱-۷ Rename by Example Values ۴۵۷
۸-۲-۲ عملگرهای تغییر نوع داده ۴۵۷
۸-۲-۲-۱ عملگرهای گسسته‌سازی ۴۵۷
۸-۲-۲-۲ عملگرهای تغییر نوع ۴۶۰
۸-۲-۳ عملگرهای تبدیل و کاهش مجموعه ویژگی‌ها ۴۶۳
۸-۲-۳-۱ بخش Generation ۴۶۳
۸-۲-۳-۲ بخش Transformation ۴۶۷
۸-۲-۳-۳ بخش Selection ۴۶۹
۸-۲-۴ عملگرهای دست‌کاری مقادیر ۴۷۶
۸-۲-۴-۱ Normalize ۴۷۶
۸-۲-۴-۲ De-Normalize ۴۷۷
۸-۲-۴-۳ Map ۴۷۷
۸-۲-۴-۴ Set data ۴۷۸
۸-۲-۵ عملگرهای پاک‌سازی داده ۴۷۸
۸-۲-۵-۱ عملگرهای شناسایی داده‌های پرت ۴۷۸
۸-۲-۵-۲ عملگرهای جایگزینی مقادیر از دست رفته ۴۷۹
۸-۲-۶ عملگرهای فیلترینگ ۴۸۰
۸-۲-۶-۱ عملگرهای نمونه‌برداری ۴۸۰
۸-۲-۶-۲ سایر عملگرهای فیلترینگ ۴۸۳

Telegram: @Farabar_BI

فرابر
www.farabar.net

۸-۲-۷ عملگرهای مرتب‌سازی داده‌ها ۴۸۵
۸-۲-۷-۱ Sort ۴۸۵
۸-۲-۷-۲ Sort by Pareto Rank ۴۸۵
۸-۲-۷-۳ Shuffle ۴۸۵
۸-۲-۸ عملگرهای انبوهش داده ۴۸۶
۸-۲-۸-۱ Aggregate ۴۸۶
۸-۲-۹ عملگرهای عملیات مجموعه‌ای ۴۸۷
۸-۲-۹-۱ Append ۴۸۷
۸-۲-۹-۲ Join ۴۸۸
۸-۲-۹-۳ Set Minus ۴۸۸
۸-۲-۹-۴ Intersect ۴۸۸
۸-۲-۹-۵ Union ۴۸۹
۸-۲-۹-۶ Superset ۴۸۹
۸-۲-۹-۷ Cartesian Product ۴۹۰
۸-۳ عملگرهای مدل‌سازی در نرم‌افزار RapidMiner ۴۹۰
۸-۳-۱ دسته‌بندی و رگرسیون ۴۹۱
۸-۳-۱-۱ استنتاج‌های درختی ۴۹۱
۸-۳-۱-۲ استنتاج مبتنی بر قوانین ۴۹۵
۸-۳-۱-۳ روش‌های مدل‌سازی Lazy ۴۹۷
۸-۳-۱-۴ مدل‌سازی Bayesian ۴۹۸
۸-۳-۱-۵ آموزش دسته‌بندهای شبکه‌های عصبی مصنوعی ۴۹۹
۸-۳-۱-۶ مدل‌سازی ماشین بردار پشتیبان ۵۰۲
۸-۳-۱-۷ مدل‌سازی Meta ۵۰۴
۸-۳-۱-۸ عملگرهای رگرسیون ۵۰۶
۸-۳-۲ وزن‌دهی ویژگی‌ها ۵۰۹
۸-۳-۲-۱ Weight by Information Gain ۵۰۹
۸-۳-۲-۲ Weight by Rule ۵۰۹
۸-۳-۲-۳ Weight by Value Average ۵۱۰

Telegram: @Farabar_BI

فرابر
www.farabar.net

۸-۲-۳-۴ Weight by Deviation ۵۱۰
۸-۲-۳-۵ Weight by Chi Squared Statistic ۵۱۱
۸-۲-۳-۶ Weight by Tree Importance ۵۱۱
۸-۲-۳-۷ Weight by User Specification ۵۱۱
۸-۲-۳-۸ Data to Weight ۵۱۲
۸-۲-۳-۹ Weight to data ۵۱۲
۸-۳-۳ خوشه‌بندی و قطعه‌بندی ۵۱۲
۸-۳-۳-۱ K-Means ۵۱۳
۸-۳-۳-۲ K-Means (Kernel) ۵۱۳
۸-۳-۳-۳ K-Medoids ۵۱۳
۸-۳-۴ کاوش مجموعه اشیاء مکرر ۵۱۴
۸-۳-۴-۱ FP-Growth ۵۱۴
۸-۳-۴-۲ Create Association rules ۵۱۴
۸-۳-۵ عملگرهای اعمال مدل روی داده ۵۱۵
۸-۳-۵-۱ Apply Model ۵۱۵
۸-۴ عملگرهای ارزیابی و تفسیر مدل در نرم‌افزار RapidMiner ۵۱۵
۸-۴-۱ الگوریتم‌های ارزیابی دسته‌بندی ۵۱۵
۸-۴-۱-۱ Split Validation ۵۱۶
۸-۴-۱-۲ X-validation ۵۱۶
۸-۴-۱-۳ Bootstrapping validation ۵۱۷
۸-۴-۲ عملگرهای ارزیابی الگوریتم‌های خوشه‌بندی ۵۱۸
۸-۴-۲-۱ Map Clustering on Label ۵۱۸
۸-۴-۲-۲ cluster distance performance ۵۱۸
۸-۴-۳ سایر عملگرهای ارزیابی و تفسیر مدل ۵۱۹
۸-۴-۳-۱ Performance ۵۱۹
۸-۴-۳-۲ Performance(User-Based) ۵۱۹
۸-۴-۳-۳ Performance classification ۵۲۰
۸-۵ خلاصه ۵۲۰