

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

داده کاوی باریدمانیر - امتیاز دہی، اعتبار سنجی و ابزار

تألیف:

شہاب طارقیان

برہام سوری

محمد صدیقی منش

علی صدیقی منش

مجتبی احمدی

سال ۱۳۹۸

سرشناسه: صدیقی منش، محمد

عنوان و نام پدید آور: داده کاوی با رپیدماینر - امتیاز دهی، اعتبارسنجی و ابزار - تألیف: محمد صدیقی منش - برسام سوری - شهاب طارقیان - مجتبی احمدی - علی صدیقی منش

مشخصات نشر: قم، الهام نور، ۱۳۹۸.

مشخصات ظاهری: ۳۰۹ ص، مصور، نمودار

شابک: ۹۷۸-۶۰۰-۶۹۴۷-۶۱-۷ / ۴۰۰۰۰ ریال

وضعیت فهرست نویسی: فیا.

موضوع: رپید ماینر (منبع الکترونیکی)، داده کاوی، نرم افزار

موضوع: RapidMiner (Electronic resource)-Data Mining- Software

شابنامه افزوده: صدیقی منش، محمد، ۱۳۶۶-

رده بندی کنگره: OA۷۶/۹

رده بندی دیویی: ۰۰۶/۳۱۲

شماره کتابشناسی ملی: ۶۰۴۶۶۰۱

Contact:

Email: Ali.sedighimanesh@gmail.com, 09193508510

Email: Mohammad.sedighimanesh@gmail.com, 09193523413

انتشارات الهام نور

نام کتاب: داده کاوی با رپیدماینر - امتیاز دهی، اعتبارسنجی و ابزار

تألیف: محمد صدیقی منش، برسام سوری، شهاب طارقیان، مجتبی احمدی، علی صدیقی منش

تیراژ / چاپخانه: ۳۰۰ نسخه / شریعت

قطع / تعداد صفحات / صحافی: وزیری / ۳۰۹ صفحه / براتی

نوبت چاپ / قیمت: اول / ۱۳۹۸ / ۴۰۰۰۰ تومان

شابک: ۹۷۸-۶۰۰-۶۹۴۷-۶۱-۷

کلیه حقوق محفوظ و مخصوص مؤلفین می باشد.

فهرست مطالب

۱۴.....	فصل پنجم - امتیازدهی
۱۴.....	Apply Model
۱۷.....	Explain Predictions
۲۰.....	Model Simulator
۲۲.....	Prescriptive Analytics
۲۶.....	۵-۱- اطمینان‌ها
۲۶.....	Apply Threshold
۲۸.....	Create Threshold
۳۱.....	Drop Uncertain Predictions
۳۴.....	Find Threshold
۳۸.....	فصل ششم - اعتبارسنجی
۳۸.....	Bootstrapping Validation
۴۲.....	Cross Validation
۵۱.....	Split Validation
۵۵.....	Wrapper Split Validation
۵۷.....	Wrapper-X-Validation
۶۱.....	۶-۱ عملکرد
۶۱.....	Combine Performances
۶۳.....	Extract Performance
۶۶.....	Performance

۷۰	Performance (Min-Max)
۷۱	Performance to Data
۷۳	۴-۱-۱ پیشگویانه
۹۷	۴-۱-۲ قطعه‌بندی
۱۰۹	۳-۱-۴ آزمون‌های اهمیت
۱۱۵	۴-۲ تصویری
۱۱۵	Compare ROCs
۱۱۹	Create Lift Chart
۱۲۲	Lift Chart (Simple)
۱۲۴	Visualize Model by SOM
۱۲۶	فصل هفتم - ابزار
۱۲۶	Execute Process
۱۲۹	Multiply
۱۳۱	Schedule Process
۱۳۳	Subprocess
۱۳۶	۷-۱-۱ اسکریپت نویسی
۱۳۶	Execute Program
۱۴۰	Execute Python
۱۴۵	Execute R
۱۴۸	Execute SQL
۱۵۷	۷-۲ کنترل فرآیند

۱۵۷	Publish to App
۱۵۹	Recall
۱۶۱	Recall from App
۱۶۴	Remember
۱۶۷	۱-۲-۷ ملقه‌ها
۲۰۵	۲-۲-۷ انشعاب‌ها
۲۱۲	۳-۲-۷ مجموعه‌ها (collections)
۲۱۹	۴-۲-۷ استثناء‌ها
۲۲۳	۳-۳-۷ (مجموعه‌ها)
۲۲۳	Extract Macro
۲۳۰	Generate Macro
۲۳۳	Set Macro
۲۳۷	Set Macros
۲۴۰	۴-۷ فایل‌ها
۲۴۰	Add Entry to Archive File
۲۴۳	Copy File
۲۴۵	Create Archive File
۲۴۷	Create Directory
۲۴۹	Delete File
۲۵۱	Move File
۲۵۳	Rename File

۲۵۵.....	Write Message
۲۵۷.....	Write as Text
۲۵۹.....	۷-۵- ماشینها.....
۲۵۹.....	Annotate
۲۶۱.....	Annotations to Data
۲۶۲.....	Data to Annotations
۲۶۴.....	Extract Macro from Annotation
۲۶۶.....	۷-۶- لایه کردن.....
۲۶۶.....	Extract Log Value
۲۶۸.....	Log
۲۷۴.....	Log to Data
۲۷۷.....	Provide Macro as Log Value
۲۸۰.....	۷-۷- گمنام سازی دادهها.....
۲۸۰.....	De-Obfuscate
۲۸۱.....	Obfuscate
۲۸۲.....	۷-۸- تولید دادههای تصادفی.....
۲۸۲.....	Add Noise
۲۸۷.....	Generate Data
۲۹۲.....	Generate Direct Mailing Data
۲۹۴.....	Generate Multi-Label Data
۲۹۶.....	Generate Nominal Data

٢٩٨.....Generate Sales Data

٣٠٠.....Generate Transaction Data

٣٠٢.....٧-٩ - متفرقة

٣٠٣.....Free Memory

٣٠٤.....Join Paths

٣٠٥.....Materialize Data

٣٠٨.....Register Visualization from Database

فهرست اشکال

- شکل (۵-۱): فرآیند آموزشی "آموزش دادن و بکار بردن یک مدل رگرسیون خطی". ۱۵
- شکل (۵-۲): فرآیند آموزشی "اپراتور Explain Predictions برای Titanic". ۱۹
- شکل (۵-۳): فرآیند آموزشی "اپراتور Model Simulator برای داده‌های Titanic". ۲۲
- شکل (۵-۴): فرآیند آموزشی "اپراتور Prescriptive Analytics برای داده‌های Titanic". ۲۵
- شکل (۵-۵): فرآیند آموزشی "ایجاد و بکار بردن آستانه‌ها". ۲۷
- شکل (۵-۶): فرآیند آموزشی "ایجاد و بکار بردن آستانه‌ها". ۳۰
- شکل (۵-۷): فرآیند آموزشی "مدف پیش‌بینی‌های نامشخص از اپراتور Naive Bayes". ۳۳
- شکل (۵-۸): فرآیند آموزشی "مقدمه‌ای بر اپراتور Find Threshold". ۳۷
- شکل (۶-۱): فرآیند آموزشی "اعتبارسنجی مدل‌ها با استفاده از Bootstrapping Validation". ۴۱
- شکل (۶-۲): فرآیند آموزشی "علت اعتبارسنجی مدل‌ها". ۴۶
- شکل (۶-۳): فرآیند آموزشی "اعتبارسنجی مدل‌ها با استفاده از Cross Validation". ۴۷
- شکل (۶-۴): فرآیند آموزشی "فرستادن نتایج از فرآیند آموزشی به زیرفرآیند آزمایشی با کمک پورت‌های through". ۴۹
- شکل (۶-۵): فرآیند آموزشی "استفاده از صفت batch برای تقسیم کردن داده‌های آموزشی". ۵۰
- شکل (۶-۶): فرآیند آموزشی "اعتبارسنجی مدل‌ها با استفاده از Split Validation". ۵۵
- شکل (۶-۷): فرآیند آموزشی "ارزیابی یک الگوی انتقاب صفات". ۶۰
- شکل (۶-۸): فرآیند آموزشی "مقدمه‌ای بر اپراتور Combine Performances". ۶۲
- شکل (۶-۹): فرآیند آموزشی "مقدمه‌ای بر اپراتور Extract Performance". ۶۵
- شکل (۶-۱۰): فرآیند آموزشی "ارزیابی عملکرد یک پیش‌بینی". ۶۹
- شکل (۶-۱۱): فرآیند آموزشی "مقدمه‌ای بر اپراتور Performance (Min-Max)". ۷۱
- شکل (۶-۱۲): فرآیند آموزشی "ارزیابی عملکرد یک پیش‌بینی". ۷۲
- شکل (۶-۱۳): فرآیند آموزشی "تولید یک بردار عملکرد با معیار «تعداد صفات»". ۷۴

شکل (۶-۱۴): فرآیند آموزشی "مداسازی معادن از صخره‌ها".....	۷۹
شکل (۶-۱۵): فرآیند آموزشی "مداسازی معادن از صخره‌ها، با Cross Validation".....	۸۰
شکل (۶-۱۶): فرآیند آموزشی "استفاده از پورت عملکرد در اپراتور Performance (Classification)".....	۸۶
شکل (۶-۱۷): فرآیند آموزشی "اندازه‌گیری هزینه‌های طبقه‌بندی نادرست یک طبقه‌بندی کننده".....	۸۹
شکل (۶-۱۸): فرآیند آموزشی "اعمال کردن اپراتور Performance (Ranking) روی مجموعه داده Golf".....	۹۲
شکل (۶-۱۹): فرآیند آموزشی "اعمال کردن اپراتور Performance (Regression) روی مجموعه داده Polynomial".....	۹۶
شکل (۶-۲۰): فرآیند آموزشی "تولید یک بردار عملکرد با معیار «Number of clusters»".....	۹۸
شکل (۶-۲۱): فرآیند آموزشی "ارزیابی عملکرد مدل فوشه‌بندی K-Means".....	۱۰۱
شکل (۶-۲۲): فرآیند آموزشی "ارزیابی عملکرد مدل فوشه‌بندی K-Medoids".....	۱۰۳
شکل (۶-۲۳): فرآیند آموزشی "ارزیابی عملکرد مدل فوشه‌بندی K-Medoids".....	۱۰۶
شکل (۶-۲۴): فرآیند آموزشی "مقدمه‌ای بر اپراتور Map Clustering on Labels".....	۱۰۸
شکل (۶-۲۵): فرآیند آموزشی "مقایسه بردارهای عملکرد با استفاده از آزمون‌های اهمیت آماری".....	۱۱۱
شکل (۶-۲۶): فرآیند آموزشی "مقایسه بردارهای عملکرد با استفاده از آزمون‌های اهمیت آماری".....	۱۱۴
شکل (۶-۲۷): فرآیند آموزشی "مقایسه گرافیکی طبقه‌بندی کننده‌های مختلف با کمک منحنی‌های ROC".....	۱۱۸
شکل (۶-۲۸): فرآیند آموزشی "ایجاد نمودار lift برای direct mailing data".....	۱۲۱
شکل (۶-۲۹): فرآیند آموزشی "نمودار lift برای Naive Bayes روی Titanic".....	۱۲۳
شکل (۶-۳۰): فرآیند آموزشی "به تصویر درآوردن درخت تصمیم توسط SOM".....	۱۲۵
شکل (۷-۱): فرآیند آموزشی "فرآیندی که بعنوان یک فرآیند جاسازی شده در فرآیند نمونه دوم بکار می‌رود".....	۱۲۸
شکل (۷-۲): فرآیند آموزشی "اجرای یک فرآیند جاسازی شده".....	۱۲۹
شکل (۷-۳): فرآیند آموزشی "تکثیر یک ExampleSet".....	۱۳۰
شکل (۷-۴): فرآیند آموزشی "استفاده از زیرفرآیندها برای سازمان‌دهی یک فرآیند".....	۱۳۵
شکل (۷-۵): فرآیند آموزشی "مقدمه‌ای بر اپراتور Execute Program".....	۱۳۸
شکل (۷-۶): فرآیند آموزشی "باز کردن Internet Explorer با کمک اپراتور Execute Program".....	۱۳۹

- شکل (۷-۷): فرآیند آموزشی "پایپ کردن داده‌ها در میان یک فرمان پوسته". ۱۳۹.....
- شکل (۷-۸): فرآیند آموزشی "فوشه‌بندی با استفاده از پایتون". ۱۴۱.....
- شکل (۷-۹): فرآیند آموزشی "سافتن و بکار بردن یک مدل با استفاده از پایتون". ۱۴۲.....
- شکل (۷-۱۰): فرآیند آموزشی "ایجاد یک نمودار با استفاده از پایتون و ذخیره کردن آن در مخزن خود". ۱۴۳.....
- شکل (۷-۱۱): فرآیند آموزشی "خواندن یک مجموعه مثال از یک فایل با استفاده از پایتون". ۱۴۴.....
- شکل (۷-۱۲): فرآیند آموزشی "مدیریت فراداده‌ها". ۱۴۴.....
- شکل (۷-۱۳): فرآیند آموزشی "آموزش دادن و بکار بردن یک مدل قطی در R". ۱۴۶.....
- شکل (۷-۱۴): فرآیند آموزشی "تولید توابع یکالی احتمال برای توابع احتمال مختلف در R". ۱۴۷.....
- شکل (۷-۱۵): فرآیند آموزشی "خواندن یک مجموعه مثال از یک فایل با استفاده از R". ۱۴۷.....
- شکل (۷-۱۶): فرآیند آموزشی "مدیریت فراداده‌ها". ۱۴۸.....
- شکل (۷-۱۷): فرآیند آموزشی "ایجاد یک جدول جدید در پایگاه داده MySQL". ۱۵۱.....
- شکل (۷-۱۸): فرآیند آموزشی "مرکت در میان صفات برای تبدیل نام صفات به مروف کوچک". ۱۵۴.....
- شکل (۷-۱۹): فرآیند آموزشی "مرکت در میان تمامی مثال‌ها برای تبدیل مقدار صفات به مروف بزرگ". ۱۵۵.....
- شکل (۷-۲۰): فرآیند آموزشی "کم کردن میانگین صفات عددی از مقدار صفات". ۱۵۶.....
- شکل (۷-۲۱): فرآیند آموزشی "مقدمه‌ای بر اپراتورهای Recall from App و Publish to App". ۱۵۹.....
- شکل (۷-۲۲): فرآیند آموزشی "مقدمه‌ای بر اپراتورهای Recall و Remember". ۱۶۱.....
- شکل (۷-۲۳): فرآیند آموزشی "مقدمه‌ای بر اپراتورهای Recall from App و Publish to App". ۱۶۴.....
- شکل (۷-۲۴): فرآیند آموزشی "مقدمه‌ای بر اپراتورهای Recall و Remember". ۱۶۶.....
- شکل (۷-۲۵): فرآیند آموزشی "استفاده از یک ملقه اصلی". ۱۶۸.....
- شکل (۷-۲۶): فرآیند آموزشی "استفاده مجدد از نتایج در یک ملقه". ۱۶۸.....
- شکل (۷-۲۷): فرآیند آموزشی "مقدمه‌ای بر اپراتور Loop Attribute Subsets". ۱۷۰.....
- شکل (۷-۲۸): فرآیند آموزشی "تولید صفات جدید در اپراتور Loop Attributes". ۱۷۵.....
- شکل (۷-۲۹): فرآیند آموزشی "مقدمه‌ای بر اپراتور Loop Batches". ۱۷۷.....
- شکل (۷-۳۰): فرآیند آموزشی "مقدمه‌ای بر اپراتور Loop Clusters". ۱۷۹.....
- شکل (۷-۳۱): فرآیند آموزشی "مقدمه‌ای بر مجموعه‌ها". ۱۸۲.....
- شکل (۷-۳۲): فرآیند آموزشی "انتخاب ExampleSet‌ی که بهترین عملکرد را دارد". ۱۸۴.....

شکل (۷-۳۳):	فرآیند آموزشی "کم کردن میانگین یک صفت از همه مثال‌ها".	۱۸۷
شکل (۷-۳۴):	فرآیند آموزشی "تولید یک ExampleSet با نام همه فایل‌های یک پوشه".	۱۹۰
شکل (۷-۳۵):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Loop Labels".	۱۹۲
شکل (۷-۳۶):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Loop Labels".	۱۹۵
شکل (۷-۳۷):	فرآیند آموزشی "استفاده از اپراتور Loop Values در عملیات پیش‌پردازشی پیچیده".	۲۰۰
شکل (۷-۳۸):	فرآیند آموزشی "میانگین‌گیری از بردارهای عملکرد".	۲۰۲
شکل (۷-۳۹):	فرآیند آموزشی "تمویل دادن بهترین بردار عملکرد".	۲۰۴
شکل (۷-۴۰):	فرآیند آموزشی "بکار بردن زیرفرآیندهای مختلف روی مجموعه داده Golf بر اساس مقدار عملکرد".	۲۰۸
شکل (۷-۴۱):	فرآیند آموزشی "بکار بردن اپراتورهای طبقه‌بندی مختلف روی مجموعه داده Golf با استفاده از اپراتور Select Subprocess".	۲۱۱
شکل (۷-۴۲):	فرآیند آموزشی "مقدمه‌ای بر مجموعه‌ها".	۲۱۴
شکل (۷-۴۳):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Flatten Collection".	۲۱۶
شکل (۷-۴۴):	فرآیند آموزشی "مقدمه‌ای بر مجموعه‌ها".	۲۱۸
شکل (۷-۴۵):	فرآیند آموزشی "استفاده از Try/Catch برای مدیریت استثناء‌ها".	۲۲۰
شکل (۷-۴۶):	فرآیند آموزشی "اگر هیچ مثالی رد نشود، استثناء مذف شود".	۲۲۲
شکل (۷-۴۷):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Extract Macro".	۲۲۶
شکل (۷-۴۸):	فرآیند آموزشی "تعریف دوباره یک ماکرو با استفاده از اپراتور Extract Macro".	۲۲۷
شکل (۷-۴۹):	فرآیند آموزشی "استفاده از اپراتور Extract Macro در عملیات پیش‌پردازشی پیچیده".	۲۲۹
شکل (۷-۵۰):	فرآیند آموزشی "تولید یک ماکروی جدید از روی یک ماکروی موجود".	۲۳۲
شکل (۷-۵۱):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Set Macro".	۲۳۶
شکل (۷-۵۲):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Set Macros".	۲۳۹
شکل (۷-۵۳):	فرآیند آموزشی "ایجاد و ذخیره کردن یک فایل zip".	۲۴۲
شکل (۷-۵۴):	فرآیند آموزشی "ذخیره کردن داده‌هایی که جدیداً ایجاد شده در یک فایل zip".	۲۴۳
شکل (۷-۵۵):	فرآیند آموزشی "نوشتن مجموعه داده Labor-Negotiations در یک فایل excel و کپی کردن آن".	۲۴۴

شکل (۷-۵۶):	فرآیند آموزشی "ایجاد و ذخیره کردن یک فایل zip".	۲۴۶
شکل (۷-۵۷):	فرآیند آموزشی "ایجاد یک پوشه جدید در درایو D".	۲۴۸
شکل (۷-۵۸):	فرآیند آموزشی "نوشتن مجموعه داده Labor-Negotiations در یک فایل excel و کپی کردن آن".	۲۵۰
شکل (۷-۵۹):	فرآیند آموزشی "نوشتن مجموعه داده Labor-Negotiations در یک فایل excel و جایا کردن آن".	۲۵۲
شکل (۷-۶۰):	فرآیند آموزشی "نوشتن مجموعه داده Labor-Negotiations در یک فایل excel و تغییر دادن نام آن".	۲۵۴
شکل (۷-۶۱):	فرآیند آموزشی "استفاده از Write Message بعنوان یک لاک کننده سفارشی".	۲۵۶
شکل (۷-۶۲):	فرآیند آموزشی "نوشتن چندین نتیجه در یک فایل".	۲۵۸
شکل (۷-۶۳):	فرآیند آموزشی "ماشین‌نویسی بر یک مجموعه داده".	۲۶۰
شکل (۷-۶۴):	فرآیند آموزشی "استخراج ماشینیه‌ها و قرار دادن آنها در یک مجموعه داده".	۲۶۲
شکل (۷-۶۵):	فرآیند آموزشی "ماشین‌نویسی بر یک شیء از روی یک مجموعه داده".	۲۶۳
شکل (۷-۶۶):	فرآیند آموزشی "استخراج ماشینیه Source از یک مجموعه مثال".	۲۶۵
شکل (۷-۶۷):	فرآیند آموزشی "لاک کردن مقدار یک صفت با کمک اپراتور Extract Log Value".	۲۶۷
شکل (۷-۶۸):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Log".	۲۷۲
شکل (۷-۶۹):	فرآیند آموزشی "مشاهده فضای آموزش در برابر فضای آزمایش با کمک اپراتور Log".	۲۷۳
شکل (۷-۷۰):	فرآیند آموزشی "دسترسی به فضای آموزش در برابر فضای آزمایش با کمک اپراتورهای Log و Log to Data".	۲۷۶
شکل (۷-۷۱):	فرآیند آموزشی "لاک کردن مقدار یک ماکرو با یا بدون اپراتور Provide Macro as Log Value".	۲۷۹
شکل (۷-۷۲):	فرآیند آموزشی "افزودن نویز به مجموعه داده Polynomial".	۲۸۶
شکل (۷-۷۳):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Generate Data".	۲۹۱
شکل (۷-۷۴):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Generate Direct Mailing Data".	۲۹۳
شکل (۷-۷۵):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Generate Multi-Label Data".	۲۹۵
شکل (۷-۷۶):	فرآیند آموزشی "مقدمه‌ای بر اپراتور Generate Nominal Data".	۲۹۷

- شکل(۷-۷۷): فرآیند آموزشی "مقدمه‌ای بر اپراتور Generate Sales Data".....۲۹۹
- شکل(۷-۷۸): فرآیند آموزشی "مقدمه‌ای بر اپراتور Generate Transaction Data".....۳۰۱
- شکل(۷-۷۹): فرآیند آموزشی "مقدمه‌ای بر اپراتور Free Memory".....۳۰۳
- شکل(۷-۸۰): فرآیند آموزشی "برگرداندن اولین شیء غیر تهی".....۳۰۵
- شکل(۷-۸۱): فرآیند آموزشی "ایجاد یک رونوشت تازه از یک ExampleSet".....۳۰۷