

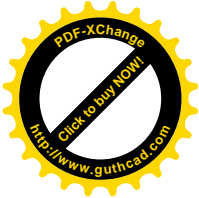
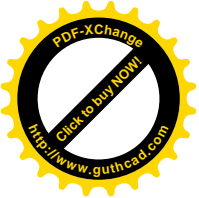
به نام خدا

آموزش گام به گام داده کاوی با RapidMiner

مؤلف:

دکتر مهدی اسماعیلی

(عضویت علمی دانشگاه)



شابک	: 978-600-6004-81-5
شماره کتابشناسی ملی	: ۳۲۹۷۵۴۴
عنوان و نام پدیدآور	: آموزش گام به گام داده‌کاوی با RapidMiner / مؤلف مهدی اسماعیلی.
مشخصات نشر	: تهران: آتی نگر، ۱۳۹۲.
مشخصات ظاهری	: ۱۳۶ص: مصور، جدول، نمودار.
یادداشت	: واژه‌نامه.
موضوع	: ریید ماینر (منبع الکترونیکی)
موضوع	: داده‌کاوی -- نرم‌افزار
رده بندی دیویی	: ۰۰۶/۳۱۲
رده بندی کنگره	: QA۹۷۶/د۲ الف ۱۴۷ ۱۳۹۲
سرشناسه	: اسماعیلی، مهدی، ۱۳۵۰ -
وضعیت فهرست نویسی	: فیبا



انتشارات آتی نگر

عنوان: آموزش گام به گام داده‌کاوی با RapidMiner

مؤلف: دکتر مهدی اسماعیلی

ناشر: آتی نگر

تیراژ: ۱۰۰۰

طراحی جلد، صفحه‌آرایی: همتا بیداریان

چاپ اول: ۱۳۹۲

قیمت: ۹۰,۰۰۰ ریال (به همراه CD)

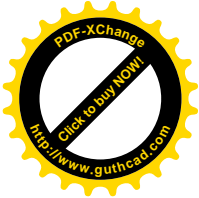
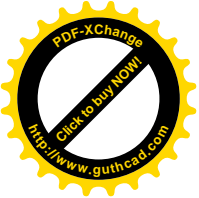
شابک: ۹۷۸-۶۰۰-۶۰۰۴-۸۱-۵

تلفن مرکز پخش: ۰۹۳۵۷۱۷۱۷۱۳ - ۶۶۵۶۵۳۳۶-۸

آدرس: تهران- خیابان جمالزاده جنوبی- روبه‌روی کوچه رشتچی- پلاک ۱۴۴- واحد ۲

www.ati-negar.com

(هرگونه کپی و نسخه‌برداری از مطالب این کتاب ممنوع می‌باشد).



فهرست مطالب

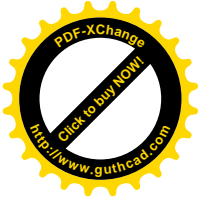
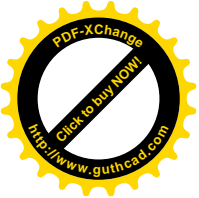
پیشگفتار

فصل اول: مقدمه RapidMiner

- ۱-۱ پیاده‌سازی و اجرای پروژه‌های داده‌کاوی ۱۰
- ۱-۲ مفاهیم اولیه ۱۱
- ۱-۳ ابزارهای داده‌کاوی ۱۴
- ۱-۴ سازماندهی کتاب ۱۶

فصل دوم: آشنایی با RapidMiner

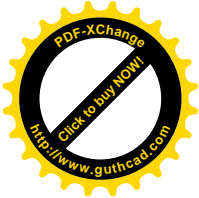
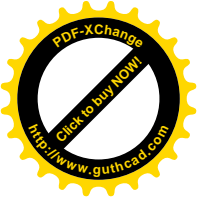
- ۲-۱ شروع کار با RapidMiner ۱۹
- ۲-۲ پرسپکتیو و نما ۲۰
- ۲-۳ پرسپکتیو خوشامدگویی ۲۲
- ۲-۴ پرسپکتیو طراحی ۲۳
- نمای عملگرها ۲۴
- نمای مخازن ۲۷
- نمای پردازش ۲۸
- عملگر ۲۸
- اتصال و ترتیب اجرای عملگرها ۳۱
- نوار ابزار نمای پردازش ۳۴
- نمای پارامترها ۳۵
- نمای مشکلات و ثبت ۳۶
- نمای راهنما ۳۷
- نمای دید اجمالی ۳۸



۳۹	۲-۵ اجرای پردازش و پرسپکتیو نتایج
۴۱	۲-۶ ایجاد یک پردازش جدید

فصل سوم: پارامترها و عملگرهای پر کاربرد

۴۳	۳-۱ پارامترهای پر کاربرد
۴۳	پارامتر attribute filter type
۴۶	پارامتر invert selection
۴۶	پارامتر include special attributes
۴۶	پارامتر create view
۴۶	پارامتر use underscore in name
۴۶	پارامتر use standard constants
۴۷	پارامتر use local random seed
۴۷	پارامترهای موازیسازی
۴۷	۳-۲ عملگرهای کنترلی
۴۷	عملگر Subprocess
۴۸	عملگر Multiply
۴۸	عملگرهای Branch و Select Subprocess
۴۹	عملگرهای Recall و Remember
۴۹	عملگر Join Paths
۴۹	عملگر Handle Exception
۴۹	عملگرهای مدیریت پارامترها
۵۲	عملگرهای مدیریت ماکروها
۵۳	عملگرهای مربوط به تولید داده‌های تصادفی
۵۴	عملگر Set Role
۵۴	عملگرهای Free Memory و Materialize Data
۵۴	عملگرهای مدیریت حلقه تکرار
۵۶	عملگر Sort
۵۶	عملگر Aggregate



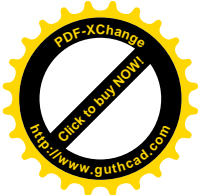
فصل چهارم: آماده‌سازی داده‌ها ۵۹

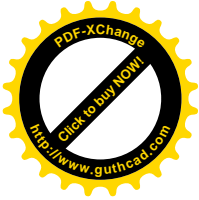
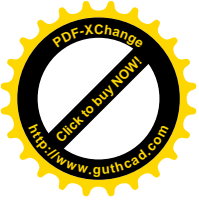
۴-۱	ورود و بارگذاری داده‌ها	۶۰
۴-۲	ایجاد صفت‌خاصه جدید و انتخاب صفات خاصه	۶۵
۴-۳	نمونه‌گیری	۶۹
۴-۴	مقادیر مفقود شده	۷۱
۴-۵	تغییر نوع داده‌ها و نرمال‌سازی	۷۳
۴-۶	گسسته‌سازی	۷۸
۴-۷	کاهش ابعاد داده‌ها	۸۱
۴-۸	کاهش ابعاد از طریق تصویر کردن	۸۴
۴-۹	داده‌های پرت	۸۶
۴-۱۰	عملیات مجموعه‌ها	۸۸

فصل پنجم: مدل‌سازی ۹۱

۵-۱	دسته‌بندی	۹۲
	درخت تصمیم	۹۳
	رگرسیون	۱۰۳
	بهبود کارایی تکنیک‌های دسته‌بندی	۱۱۰
	دو عملگر سودمند برای دسته‌بندی و رگرسیون	۱۱۶
۵-۲	خوشه‌بندی	۱۱۷
۵-۳	قوانین انجمنی	۱۲۱
۵-۴	ابزارهای دیداری	۱۲۷

ضمیمه الف: محتویات CD ۱۳۳

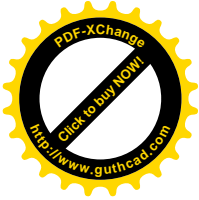




پیشگفتار

زاهد ظاهرپرست از حال ما آگاه نیست
در حق ما هر چه گوید جای هیچ اکراه نیست
بنده پیر خراباتم که لطفش دائم است
ورنه لطف شیخ و زاهد گاه هست و گاه نیست

پیشرفت‌های بوجود آمده در جمع آوری داده‌ها در طی دهه‌های اخیر باعث شده تا با حجم بزرگی از داده‌ها روبرو شویم. داده‌کاوی کوششی برای بدست آوردن اطلاعات سودمند از میان این داده‌های حجیم است و رشد بی رویه داده‌ها در سطح جهان اهمیت داده‌کاوی را دوچندان کرده است. بدیهی است برای شروع پروژه‌های داده‌کاوی نیازمند یک نرم‌افزار هستیم و هر یک از نرم‌افزارها و یا ابزارهای حوزه داده‌کاوی دارای مزایا و معایب مربوط به خود هستند. در این چند سال اخیر ویژگی‌های یکی از آن‌ها به نام RapidMiner کاربران زیادی را به خود جلب کرده است. عدم نیاز به برنامه‌نویسی، تنوع الگوریتم‌های موجود در نرم‌افزار و بهنگام‌سازی این الگوریتم‌ها در بازه‌های زمانی کوتاه مدت از جمله قابلیت‌هایی است که باعث شد این ابزار در میان کاربران حوزه داده‌کاوی از محبوبیت ویژه‌ای برخوردار شود. از آنجا که ما نیز توجه به نیازهای کاربران را یکی از مولفه‌های اصلی در موفقیت یک محصول می‌دانیم، این ابزار را از میان بسیاری از نرم‌افزارها انتخاب کردیم. کتاب حاضر حاوی آموزش این نرم‌افزار منبع باز است که در پنج فصل تنظیم شده است. سعی کرده‌ایم که قبل از استفاده از هر مینو، عملگر و پارامتر توضیح مختصری نیز در مورد آن‌ها از دیدگاه داده‌کاوی داشته باشیم تا خواننده محترم کتاب دید روشنی از این امکانات و کاربرد آن نیز داشته باشد. پس از فصل آغازین کتاب که مقدمه‌ای در مورد چگونگی انجام پروژه‌های داده‌کاوی است، در فصل دوم با محیط این نرم‌افزار آشنا می‌شویم. عملگرها و پارامترهای پرکاربرد و رایج را در فصل سوم مطالعه کنید. می‌توانید با گذر کردن از این فصل و در حین مطالعه کتاب در صورت لزوم به فصل سوم رجوع کنید. شاید بهتر بود این فصل به عنوان یک ضمیمه در نظر گرفته می‌شد. تکنیک‌های آماده‌سازی داده‌ها موضوع فصل چهارم کتاب را تشکیل می‌دهد. آن‌هایی که درگیر پروژه‌های داده‌کاوی بوده‌اند خوب می‌دانند که مرحله آمایش داده‌ها از چه اهمیت بالایی برخوردار است. در فصل آخر کتاب یعنی فصل پنجم اجرای برخی از تکنیک‌های داده‌کاوی را بررسی می‌کنیم. در این فصل تنها چند روش رایج و محبوب داده‌کاوی با مثال‌های کاربردی شرح داده شده است. بدون شک قابلیت نرم‌افزار مزبور و



همچنین روش‌های داده‌کاوی به مطالبی که در این کتاب آموزشی آمده است محدود نمی‌شود و امید داریم این کتاب گام‌های نخستین سفر شما در این حوزه کاربردی باشد. در نگارش کتاب حاضر تمام سعی و تلاش خود را به کار برده‌ایم تا از پیچیده‌نویسی و مبهم‌گویی پرهیز کنیم و به سوی ساده‌نویسی حرکت کنیم. با وجود همه سعی و تلاشی که در تمام مراحل آماده‌سازی این کتاب انجام گرفته است، یقین داریم که عاری از اشتباه نیست. در اینجا لازم می‌دانم از ناشر محترم به خصوص همکار دانشورم آقای رامین مولاناپور به پاس تلاش در چاپ و نشر این کتاب صمیمانه تشکر کنم. در آخر نیز ضمن سپاسگزاری از همه کسانی که مرا یاری داده‌اند و با پذیرش مسئولیت هرگونه کاستی احتمالی، امیدوارم که این اندک مفید افتد.

مهدی اسماعیلی

مهرماه ۱۳۹۲