

داده‌کاوی و کشف دانش

گام به گام با نرم افزار

تألیف :

سمیه علیزاده و سمیرا ملک محمدی

www.ketab.ir

سرشناسه	: علیزاده، سمیه، ۱۳۵۴ -
عنوان و نام پدیدآور	: داده‌کاوی و کشف دانش گام به گام با نرم‌افزار Clementine / تالیف سمیه علیزاده، سمیرا ملک‌محمدی.
مشخصات نشر	: تهران: دانشگاه صنعتی خواجه نصیرالدین طوسی، انتشارات، ۱۳۹۰.
مشخصات ظاهری	: ۲۶۷ ص.: مصور (بخشی رنگی)، جدول، نمودار.
شابک	: ۷۵۰۰۰ ریال: 978-600-6383-18-7
وضعیت فهرست‌نویسی	: فیبا
یادداشت	: واژه‌نامه.
یادداشت	: کتابخانه: ص. [۳۲۵]
موضوع	: داده‌کاوی
موضوع	: داده‌کاوی-- نرم‌افزار
شناسه افزوده	: ملک‌محمدی، سمیرا، ۱۳۶۰ -
شناسه افزوده	: دانشگاه صنعتی خواجه نصیرالدین طوسی
رده بندی کنگره	: ۱۳۹۰ QA ۷۶/۹/۸۴۲۵
رده بندی دیویی	: ۰۰۴/۳۱۱۳
شماره کتابشناسی ملی	: ۲۶۹۸۰۷۸

نام کتاب: داده‌کاوی و کشف دانش گام به گام با نرم افزار Clementine

مولف: دکتر سمیه علیزاده عضو هیئت علمی دانشکده مهندسی صنایع، دانشگاه صنعتی

خواجه نصیرالدین طوسی، سمیرا ملک محمدی

ناشر: انتشارات دانشگاه صنعتی خواجه نصیرالدین طوسی

نوبت چاپ: اول

تاریخ چاپ: اسفندماه ۱۳۹۰

تیراژ: ۱۰۰۰ جلد

قیمت: ۷۵۰۰۰ تومان

کد کتاب: ۳۰۴

ISBN: 978- 600-6383-18-7

شابک: ۹۷۸- ۶۰۰-۶۳۸۳-۱۸-۷

چاپ و لیتوگرافی: پدید رنگ، هوررنگ

صحافی: گرنامی

آدرس و تلفن مرکز بخش و فروش: خیابان ولیعصر (عج)، بالاتر از میدان ونک، تقاطع میرداماد،

روبروی ساختمان اسکان (۸۸۷۷۲۲۷۷-۲۱)

(حق چاپ برای ناشر محفوظ است)

۱ - فصل اول

۱ مقدمه
۵ پروژه‌های Clementine
۶ ابزارهای Clementine
۷ جریان‌ها
۸ ساخت جریان داده
۹ کارکردن با گره‌ها:
۱۵ تنظیم گزینه‌های گره
۱۷ منابع (Sources)
۳۲ گره‌های خروجی (Output Nodes)

فصل دوم

۳۹ مقدمه
۴۰ گره‌هایی که مرتبط با فیلدها و عملیات آنها هستند
۴۰ معرفی گره Type
۴۱ انواع داده‌ها
۴۹ معرفی گره Filtering یا تغییر نام Rename
۵۲ معرفی گره Ensemble
۵۵ معرفی گره Derive
۶۳ معرفی گره Filler
۶۷ معرفی گره بی نام سازی Anonymize
۶۸ تنظیم گزینه‌ها برای گره Anonymize
۷۱ بی نام سازی مقادیر فیلد
۷۳ معرفی گره دسته بندی مجدد (Reclassify)
۷۶ ذخیره سازی فیلدهای Reclassify
۷۷ معرفی گره گسسته سازی (Binning)
۷۸ تنظیم گزینه‌ها برای گره Binning
۸۰ گسسته سازی با عرض ثابت
۸۱ روش کاشی کاری
۸۳ روش رتبه بندی

۸۴.....	میانگین لانحراف استاندارد
۸۵.....	روش Binning بهینه
۸۸.....	معرفی گروه RFM Analysis
۹۰.....	مقادیر دسته‌ای در تحلیل گروه RFM
۹۱.....	معرفی گروه Partition
۹۳.....	معرفی گروه Set to Flag
۹۴.....	معرفی گروه Restructure
۹۶.....	استفاده از گروه Restructure با گروه Aggregate
۹۶.....	معرفی گروه Transpose
۱۰۰.....	معرفی گروه Time Intervals
۱۰۷.....	معرفی گروه History
۱۰۸.....	تنظیم گزینه‌ها برای گروه تاریخ
۱۰۹.....	معرفی گروه Field Recorder
۱۱۲.....	عملیات کاهش بعد و انتخاب ویژگیها در پیش پردازش داده‌ها
۱۱۲.....	معرفی گروه PCA/Factor
۱۱۲.....	تعریف تحلیل مؤلفه‌های اصلی
۱۱۳.....	گزینه‌های گروه PCA/Factor
۱۱۵.....	گزینه‌های PCA/Factor Node Rotation
۱۱۶.....	معادلات PCA/Factor Model Nugget
۱۱۸.....	خلاصه PCA/Factor Model Nugget
۱۱۸.....	خروجی پیشرفته PCA/Factor Model Nugget
۱۱۹.....	معرفی گروه Feature Selection
۱۲۰.....	تنظیم گزینه مدل Feature Selection
۱۲۰.....	Screening Predictors
۱۲۲.....	نتایج مدل Feature Selection
۱۲۳.....	انتخاب فیلدها بر اساس اهمیت
۱۲۴.....	ساخت یک فیلتر از یک مدل انتخاب ویژگی
۱۲۵.....	معرفی گروه Anomaly Detection
۱۲۷.....	تنظیم گزینه‌های Anomaly Detection
۱۲۸.....	تنظیم گزینه خبره در Anomaly Detection
۱۳۱.....	جزئیات مدل Anomaly Detection
۱۳۲.....	خلاصه مدل Anomaly Detection
۱۳۳.....	تنظیم گزینه‌های مدل Anomaly Detection

۱۳۵	آماده سازی داده‌ها و پردازشهای اولیه بر روی آنها
۱۳۵	آماده سازی داده ها برای تحلیل
۱۳۵	ساخت یک Stream
۱۳۸	استفاده از Browsing Statistics and Chart
۱۴۰	مدیریت نقاط پرت
۱۴۳	مدیریت داده‌های مفقوده
۱۴۵	مدیریت مقادیر مفقوده
۱۴۸	توابع Clementine برای برخورد با مقادیر مفقوده

فصل سوم

۱۵۱	مدل‌های خوشه بندی
۱۵۳	معرفی گره Kohonen
۱۵۵	گزینه‌های Model گره Kohonen
۱۵۸	تنظیم گزینه‌های Expert گره Kohonen
۱۶۰	قطعات مدل Kohonen
۱۶۰	جزئیات خوشه بندی توسط مدل Kohonen
۱۶۲	خلاصه مدل Kohonen
۱۶۳	معرفی گره K-Means
۱۶۴	تنظیمات Model گره K-Means
۱۶۵	گزینه‌های Expert گره K-Means
۱۶۶	خروجی مدل K-Means
۱۶۷	جزئیات مدل خوشه‌بندی K-Means
۱۶۸	خلاصه مدل K-Means
۱۶۸	معرفی گره خوشه بندی TwoStep
۱۷۰	تنظیم گزینه‌های Model گره خوشه بندی TwoStep
۱۷۲	خروجی مدل خوشه بندی TwoStep
۱۷۲	جزئیات خوشه بندی مدل TwoStep
۱۷۴	خلاصه مدل TwoStep
۱۷۴	معرفی نمایشگر روشهای خوشه بندی
۱۷۵	تب Viewer خوشه
۱۷۷	درک نمایش خوشه
۱۸۱	جستجوی نمایشگر خوشه

۱۸۴	انتخاب خوشه‌ها برای نمایش
۱۸۵	انتخاب فیلدها برای نمایش
۱۸۷	مرتب سازی اقلام صفحه
۱۸۷	تنظیم گزینه‌های اهمیت
۱۸۸	نمایش خوشه‌ها به صورت متنی
۱۸۹	برای مشاهده به صورت متنی:

فصل چهارم

۱۹۱	مدل های قواعد تلازمی
۱۹۴	داده‌های جدولی در برابر داده‌های تراکنشی
۱۹۵	داده های تراکنشی
۱۹۵	داده‌های جدولی
۱۹۶	گره GRI
۱۹۷	ویژگی‌های گره GRI
۱۹۸	گره Apriori
۱۹۹	ویژگی‌های گره Apriori
۲۰۳	گره CARMA
۲۰۵	ویژگی‌های فیلدهای گره CARMA
۲۰۷	ویژگی‌های مدل گره CARMA
۲۰۸	ویژگی‌های پیشرفته گره CARMA
۲۱۰	گره Sequence
۲۱۱	ویژگی‌های فیلدهای گره Sequence
۲۱۲	ویژگی‌های مدل گره Sequence
۲۱۴	ویژگی‌های خبره گره Sequence

فصل پنجم

۲۱۷	مدل‌های درخت تصمیم
۲۱۹	الگوریتم‌های ساخت درخت
۲۲۰	ساختن درخت تعاملی
۲۲۲	رشد و هرس درخت

۲۲۳	خواندن قواعد انشعاب در بخش Viewer
۲۲۴	تعریف انشعاب‌های دلخواه
۲۲۵	دیدن جزئیات پیش بینی کننده
۲۲۶	جزئیات انشعاب و جانشینان
۲۲۷	جانشین‌ها
۲۲۷	تنظیم کردن نمایش درخت مطابق با نیاز
۲۲۸	نمایش آمارها و گراف‌ها
۲۲۹	بهره‌ها
۲۳۰	دسته‌بندی بهره‌ها
۲۳۰	آمارهای گره به گره
۲۳۱	آمارهای تجمعی
۲۳۲	چارک‌ها (Quintiles)
۲۳۳	بهره و ROI دسته بندی
۲۳۴	بهره‌های رگرسیون
۲۳۴	چارت بهره‌ها
۲۳۵	چارت ارتقا
۲۳۶	چارت پاسخ
۲۳۷	انتخاب مبتنی بر بهره‌ها
۲۳۹	بهره‌ها برای پارتیشن‌ها
۲۴۰	ریسک‌ها
۲۴۰	ذخیره کردن مدل درخت و نتایج آن
۲۴۱	تولید مدل از سازنده درخت
۲۴۲	به روز رسانی مسیر (Directives) درخت
۲۴۲	ارسال مدل ، بهره و اطلاعات ریسک
۲۴۳	تولید گره‌های فیلتر و انتخاب
۲۴۳	تولید مجموعه قوانین از درخت تصمیم
۲۴۴	گره درخت C&R
۲۴۵	هرس
۲۴۶	گزینه‌های مدل گره درخت
۲۴۷	جهت دهندگان رشد درخت
۲۴۸	ویرایش جهت دهندگان
۲۴۹	قواعد جهت دهندگان
۲۵۰	جهت دهندگان درختان CHAID

۲۵۰	استفاده از جهت دهندگان در اسکریپت‌ها
۲۵۱	گزینه‌های پیشرفته گره درخت C&R
۲۵۳	گزینه‌های Stopping گره درخت
۲۵۳	گزینه‌های Prior Probabilities
۲۵۵	گزینه‌های هزینه دسته بندی اشتباه
۲۵۶	گره CHAID
۲۵۷	گزینه‌های پیشرفته گره CHAID
۲۵۹	گره QUEST
۲۶۰	گزینه‌های پیشرفته گره QUEST
۲۶۱	گره C5.0
۲۶۳	گزینه‌های مدل گره C5.0
۲۶۵	ساخت مستقیم مدل درختی
۲۶۶	قطعه مدل درخت تصمیم
۲۶۷	امتیازدهی مدل‌های درختی
۲۶۷	کار با قطعه‌های مدل درختی
۲۶۸	قوانین مدل درخت تصمیم
۲۶۹	قوانین درخت
۲۷۰	اهمیت متغیرها
۲۷۰	سایر اطلاعات مدل
۲۷۱	Viewer مدل درخت تصمیم
۲۷۵	مدلهای Boosted C5.0

فصل ششم

۲۸۱	گره شبکه بیزین
۲۸۴	گزینه‌های مدل گره شبکه بیزین
۲۸۶	گزینه‌های خبره مدل شبکه بیزین
۲۸۹	قطعه مدل شبکه بیزین
۲۹۰	ستون چپ
۲۹۰	ستون راست
۲۹۰	تنظیمات مدل شبکه بیزین
۲۹۲	خلاصه مدل شبکه بیزین

فصل هفتم

۲۹۵	گره شبکه‌های عصبی
۲۹۷	گزینه‌های " Model " گره شبکه عصبی
۳۰۰	"گزینه‌های اضافی" گره شبکه عصبی
۳۰۲	گزینه "Expert" گره شبکه عصبی
۳۰۲	گزینه پیشرفته گره شبکه عصبی - روش "Quick"
۳۰۳	گزینه‌های پیشرفته در گره شبکه عصبی - روش multiple
۳۰۴	گزینه‌های پیشرفته در گره شبکه عصبی - روش Prune
۳۰۶	گزینه‌های پیشرفته در گره شبکه عصبی - روش RFBN
۳۰۷	سرعت یادگیری در گره شبکه عصبی

فصل هشتم

۳۰۹	دسته‌بندی با استفاده از روش ماشین پشتیبان بردار
۳۱۶	بررسی داده‌ها
۳۲۰	جمع‌بندی

فصل نهم

۳۲۱	مدل سازی سری‌های زمانی
۳۲۳	روند
۳۲۴	دوره‌های فصلی
۳۲۵	دوره‌های غیرفصلی
۳۲۶	پالسها و استپ‌ها
۳۲۷	داده‌های پرت
۳۲۹	توابع خود همبستگی و خود همبستگی جزئی
۳۳۰	تبدیلات سری
۳۳۱	سری‌های کمکی پیش بینی کننده