

داده‌کاوی با R

به همراه متن کاوی و تحلیل شبکه‌های اجتماعی

www.ketab.ir

نویسندگان :

بابک تیمورپور

حیدر نجفی

آبان ۹۴

سرشناسه	: تیمورپور، بایک، ۱۳۵۰
شناسه افزوده	: نجفی، حیدر، ۱۳۶۹
عنوان و نام پدیدآور	: داده کاوی با R به همراه متن کاوی و تحلیل شبکه‌های اجتماعی / نویسنده بایک تیمورپور، حیدر نجفی.
مشخصات ظاهری	: ۳۰۱ ص: مصور، جدول، نمودار.
مشخصات نشر	: تهران: مرکز تحقیقات و توسعه سازمان اتکا، ۱۳۹۴.
شابک	: ۱۵۰۰۰۰ ریال: ۵-۶۸-۷۶۲۴-۶۰۰-۹۷۸
وضعیت فهرست‌نویسی	: فیپا
موضوع	: داده کاوی
موضوع	: رسانه های اجتماعی
شناسه افزوده	: اتکا. مرکز تحقیقات و توسعه
رده‌بندی کنگره	: ۱۳۹۴ ت ۹/۵۲/۹۷۶ QA۷۶
رده‌بندی دیوئی	: ۰۰۶/۳۱۲
شماره کتابشناسی ملی	: ۳۹۳۳۸۳



انستیتوت مرکز تحقیقات و توسعه سازمان اتکا

عنوان:	داده کاوی با R به همراه متن کاوی و تحلیل شبکه‌های اجتماعی
مؤلفان:	بایک تیمورپور، حیدر نجفی
ویراستار:	-
زمان و نوبت چاپ:	آبان ۱۳۹۴، اول
شمارگان:	۲۰۰ نسخه
طراح جلد:	حیدر نجفی
شابک:	۵-۶۸-۷۶۲۴-۶۰۰-۹۷۸
بهاء:	۱۵۰,۰۰۰ ریال

کلیه حقوق چاپ این کتاب محفوظ و متعلق به مرکز تحقیقات، توسعه و کیفیت سازمان اتکا است. هر نوع تکثیر یا نشر تمام یا بخشی از این کتاب منوط به اجازه کتبی سازمان اتکا خواهد بود.

نشانی: تهران، خ امام خمینی (ره)، خ شهید مرادی، ساختمان شهدای مرکزی اتکا، ط ۲

تلفن: ۰۲۱-۶۱۹۱۵۱۲۴ دورنگار: ۰۲۱-۶۱۹۱۵۳۵۵ وبسایت: www.etkaec.ir

ایمیل: tec@ERDCenter.com

بسمه تعالی

«کار علمی و تحقیقاتی‌تان را جدی بگیرید... دانش خیلی از نیازهای شما در دانشگاه‌ها وجود دارد.»

مقام معظم رهبری

حضرت امام خامنه‌ای (مدظله‌العالی)

داده‌کاوی^۱ یکی از مفاهیم نوین میان‌رشته‌ای است که داده‌های حوزه‌های مختلفی همچون ریاضیات، آمار، فناوری اطلاعات، مدیریت و سایر زمینه‌های مرتبط را باهم ترکیب می‌کند تا اطلاعات و دانش مفید و موثر پنهان در حجم بزرگی از داده‌ها را استخراج نماید. امروزه با گسترش فزاینده فناوری اطلاعات، تقریباً همه سازمان‌ها حجم عظیمی از داده‌ها را در پایگاه داده خود ذخیره می‌نمایند. داده‌کاوی، این پایگاه‌ها را، متعاقب کشف و استخراج دانش، با استفاده از روشهای یادگیری ماشینی^۲ و نیمه ماشینی تحلیل می‌نماید.

در این راستا و به منظور بهره‌برداری حداکثری در حوزه داده‌کاوی، نرم‌افزارهای مختلفی برای سازمان‌ها مهیا شده که یکی از مهمترین و پرکاربردترین آنها نرم‌افزار R می‌باشد. این نرم‌افزار یک زبان برنامه‌نویسی و محیط نرم‌افزاری برای محاسبات آماری و تحلیل داده‌ها است و به‌صورت رایگان در دسترس عموم قرار گرفته است. نرم‌افزار R حاوی محدوده‌ی گسترده‌ای از تکنیک‌های آماری از جمله، مدل‌سازی خطی و غیرخطی، آزمون‌های کلاسیک آماری، تحلیل سری‌های زمانی، رده‌بندی، خوشه‌بندی و قابلیت‌های گرافیکی است.

کتاب "داده‌کاوی با نرم‌افزار R به همراه متن‌کاوی و تحلیل شبکه‌های اجتماعی" توسط پژوهشگر محترم مرکز تحقیقات و نوآوری سازمان اتکا آقای حیدر نجفی گردآوری و با همکاری دکتر تیمورپور تالیف گردیده است. ضمن سپاسگزاری از مولفین محترم، امیدوارم این اثر ارزشمند دریچه‌ای نو به فهم مسائل مرتبط به تجزیه و تحلیل داده‌ها و اطلاعات به روی مخاطبان ارجمند بگشاید و برای ارتقا سطح دانش مدیران و کارکنان این سازمان و کلیه کاربران در سطح میهن عزیز اسلامی ایران مفید واقع گردد. انشاء الله ...

محمد مهدی کربلایی

مدیرعامل سازمان اتکا

¹ Data Mining

² Machine Learning Method

فصل ۱_مقدمه‌ای بر داده‌کاوی	۱۰
۱-۱_داده‌کاوی و کشف دانش در پایگاه داده‌ها	۱۱
۲-۱_فرآیند کشف دانش	۱۳
۳-۱_عملکردهای داده‌کاوی	۱۷
۴-۱_کاربردهای داده‌کاوی	۱۸
۵-۱_انواع حوزه‌های اطلاعات	۱۹
۶-۱_انواع داده	۲۳
۷-۱_داده، اطلاعات و دانش	۲۸
۸-۱_انواع ابزارها و زبان‌های مورد استفاده در داده‌کاوی	۳۳
۹-۱_چالش‌های تحقیقات و کاربردهای داده‌کاوی	۳۴
۱۰-۱_انبار داده و پایگاه داده تراکنشی	۳۸
فصل ۲_مقدمه‌ای بر زبان R	۴۲
۱-۲_مقدمات و توابع ابتدایی	۴۳
۲-۲_مزایای منحصربه‌فرد نرم‌افزار R	۴۴
۳-۲_قراردادهای ابتدایی	۴۸
۴-۲_ساختار داده‌ها در R	۵۲
۵-۲_محاسبات ریاضی در R	۶۱
۶-۲_نوشتن توابع در R	۷۰
۷-۲_آمار توصیفی و نمودارها	۷۳
فصل ۳_تحلیل اکتشافی داده‌ها	۸۲
۱-۳_آماده‌سازی داده‌ها برای داده‌کاوی	۸۳
۲-۳_پیش‌پردازش داده‌ها	۸۳
۳-۳_تحلیل اکتشافی داده‌ها در زبان R	۹۳

فصل ۴ خوشه‌بندی.....	۱۲۲
۱-۴ خوشه‌بندی.....	۱۲۳
۲-۴ تفاوت خوشه‌بندی و دسته‌بندی.....	۱۲۳
۳-۴ فرآیند خوشه‌بندی.....	۱۲۴
۴-۴ کاربردهای خوشه‌بندی.....	۱۲۵
۵-۴ اعتبارسنجی خوشه‌ها.....	۱۲۶
۶-۴ شاخصهای اعتبارسنجی.....	۱۲۷
۷-۴ روش‌های خوشه‌بندی.....	۱۲۸
۸-۴ روش افزایش.....	۱۳۳
۹-۴ روش خوشه‌بندی سلسله‌مراتبی.....	۱۳۶
۱۰-۴ روش‌های مبتنی بر چگالی.....	۱۴۰
۱۱-۴ ارزیابی خوشه‌بندی با معیار سیلوئت.....	۱۴۱
۱۲-۴ خوشه‌بندی در نرم‌افزار R.....	۱۴۲
فصل ۵ دسته‌بندی و پیش‌بینی.....	۱۵۲
۱-۵ دسته‌بندی.....	۱۵۳
۲-۵ تفاوت دسته‌بندی و خوشه‌بندی.....	۱۵۵
۳-۵ بیز ساده.....	۱۵۵
۴-۵ دسته‌بندی بر مبنای نزدیک‌ترین همسایگی.....	۱۵۶
۵-۵ شبکه‌های عصبی.....	۱۵۷
۶-۵ درخت تصمیم.....	۱۵۸
۷-۵ پیش‌بینی.....	۱۶۰
۸-۵ قواعد انجمنی.....	۱۶۱
۹-۵ دسته‌بندی در نرم‌افزار R.....	۱۶۳
فصل ۶ سری‌های زمانی.....	۱۸۸
۱-۶ تعریف سری زمانی.....	۱۸۹
۲-۶ سری‌های زمانی در داده‌کاوی.....	۱۹۲

۱۹۴	۳-۶ اجزای سری‌های زمانی
۱۹۶	۴-۶ شناسایی، تجزیه و حذف اجزای سری‌های زمانی
۱۹۷	۵-۶ سری زمانی در نرم‌افزار R
۲۱۲	فصل ۷ تحلیل شبکه‌های اجتماعی
۲۱۳	۱-۷ تحلیل شبکه‌های اجتماعی
۲۱۷	۲-۷ انواع مرکزیت و شاخص‌های اصلی در تحلیل شبکه
۲۲۳	۳-۷ اجتماع‌یابی در شبکه
۲۲۴	۴-۷ هم‌ارزی ساختاری
۲۲۶	۵-۷ تحلیل شبکه‌ای در نرم‌افزار R
۲۶۰	فصل ۸ متن‌کاوی
۲۶۱	۱-۸ متن‌کاوی
۲۶۲	۲-۸ پیش‌پردازش متون
۲۷۱	۳-۸ خوشه‌بندی متون
۲۸۳	۴-۸ کدهای متن‌کاوی
۲۹۴	منابع