

هنر علم داده‌ها

راهنمایی برای کسانی که با داده‌ها کار می‌کنند

نویسندگان:

راجر دی-پنگ و الیزابت مانسویی

مترجمان:

دکتر محمدجعفر تارخ

دکتر شهناز پیروزفر

دکتر پریا سلیمانی



شماره ۵۱۴

سرشناسه: پنگ، راجر دی. Peng, Roger D.

عنوان و نام پدیدآور: هنر علم داده‌ها: راهنمایی برای کسانی که با داده‌ها کار می‌کنند/نویسندگان راجردی پنگ و الیزابت ماتسویی؛ مترجمان محمدجعفر تارخ، شهناز پیروزفر، پریا سلیمانی
مشخصات نشر: تهران: دانشگاه صنعتی خواجه نصیرالدین طوسی، انتشارات، ۱۴۰۱.
مشخصات ظاهری: ۱۶۳ص.

شابک: 978-622-6655-94-1

وضعیت فهرست نویسی: فیپا

یادداشت: عنوان اصلی: The Art of Data Science.

موضوع: داده‌کاوی / Data mining

شناسه افزوده: ماتسویی، الیزابت

شناسه افزوده: Matsui, Elizabeth

شناسه افزوده: تارخ، محمدجعفر. ۱۳۳۸ - مترجم

شناسه افزوده: پیروزفر، شهناز. ۱۳۵۲ - مترجم

شناسه افزوده: سلیمانی، پریا. ۱۳۵۹ - مترجم

رده‌بندی کنگره: QA۷۶/۹

رده‌بندی دیویی: ۰۰۶/۳۱۲

شماره کتابشناسی ملی: ۸۸۲۶۳۶۴

press.kntu.ac.ir



ناشر: دانشگاه صنعتی خواجه نصیرالدین طوسی

عنوان: هنر علم داده‌ها: راهنمایی برای کسانی که با داده‌ها کار می‌کنند

مؤلفان: راجردی پنگ و الیزابت ماتسویی

مترجمان: دکتر محمدجعفر تارخ (عضو هیات علمی دانشگاه صنعتی خواجه نصیرالدین طوسی)، دکتر شهناز

پیروزفر و دکتر پریا سلیمانی (اعضای هیات علمی دانشگاه آزاد اسلامی - واحد تهران جنوب)

نوبت چاپ: اول

تاریخ انتشار: اردیبهشت ۱۴۰۱

شمارگان: ۲۰۰ جلد

چاپ: نقش آفرین

صحافی: گرنامی

قیمت: ۸۵۰۰۰ تومان

تمام حقوق برای ناشر محفوظ است

خیابان میرداماد غربی - شماره ۴۷۰ - انتشارات دانشگاه صنعتی خواجه نصیرالدین طوسی - تلفن: ۸۸۸۸۱۰۵۲

میدان ونک - خیابان ولی‌عصر (ع) - بالاتر از چهارراه میرداماد - شماره ۲۶۲۶ - مرکز پخش و فروش انتشارات

تلفن: ۸۸۷۷۲۲۷۷ رایانامه: press@kntu.ac.ir - تارنما (فروش برخط): press.kntu.ac.ir

اغراق نیست که بگوییم جهان امروز، جهان داده است و ذخیره‌سازی و استفاده بهینه از داده یکی از چالش‌های اصلی سازمان‌ها در یک دهه گذشته بوده است. از این‌رو در سال‌های اخیر، تمرکز اصلی متخصصان داده و شرکت‌های تخصصی داده روی ساختن چهارچوب‌ها و راه‌حل‌هایی برای مدیریت داده‌ها (از استخراج و ذخیره‌سازی تا تحلیل و استراتژی) بوده است.

متأسفانه، فرایند تجزیه و تحلیل داده‌ها چیزی نیست که بتوانیم به صورت اثربخش توضیح دهیم. علم داده مجموعه‌ای از ابزارها، الگوریتم‌ها و اصول یادگیری ماشین است که ما قصد نداریم در این کتاب وارد جزئیات آنها شویم. هدف ما در این کتاب، نگارش فرایند تجزیه و تحلیل داده‌هاست. در واقعیت، تجزیه و تحلیل داده‌ها یک فرایند بی تکرار و غیرخطی است که در این کتاب، این فرایند به شکل یک چرخه نشان داده می‌شود که در آن یادگیری اطلاعات در هر گام صورت می‌گیرد و سپس مشخص می‌شود که چگونه به گام بعدی پیشروی می‌شود. پالایش شده و دوباره تکرار می‌شوند و یا چگونه به گام بعدی پیشروی می‌شود.

امروزه سازمان‌های زیادی علم داده را در اولویت کار خود قرار داده‌اند و سرمایه‌گذاری‌های زیادی روی آن انجام داده‌اند، مدیران این سازمان‌ها، این علم یا هنر را یک راه‌حل استراتژیک برای موفقیت کسب و کارشان می‌دانند و به همین دلیل سرمایه‌گذاری‌های زیادی روی آن انجام داده‌اند.

امید است مطالعه این کتاب، علاقمندان علم داده را با فرایند تجزیه و تحلیل داده‌ها آشنا کند.

۷	۱- هنر تجزیه و تحلیل داده ها
۱۱	۲- چرخه تجزیه و تحلیل
۱۲	۱- ۲-۱ مقدمات تجزیه و تحلیل داده ها
۱۳	۲-۲ چرخه تجزیه و تحلیل
۱۵	۲-۳ تنظیم انتظارات
۱۶	۲-۴ گردآوری اطلاعات
۱۷	۲-۵ مقایسه انتظارات با داده ها
۱۸	۲-۶ به کارگیری فرایند چرخه تجزیه و تحلیل
۲۵	۳- تشریح و پالایش پرسش
۲۹	۳-۱ انواع پرسش
۲۹	۳-۲ کاربرد چرخه تجزیه و تحلیل برای تشریح و پالایش پرسش
۲۹	۳-۳ ویژگی های یک پرسش خوب
۳۲	۳-۴ تبدیل پرسش به مسأله داده ای
۳۵	۳-۵ مطالعه موردی
۳۹	۳-۶ نتیجه گیری
۴۱	۴- تجزیه و تحلیل داده های اکتشافی
۴۳	۴-۱ چک لیست تجزیه و تحلیل داده های اکتشافی یک مطالعه موردی
۴۴	۴-۲ پرسش تان را مشخص کنید
۴۵	۴-۳ اطلاعات خود را مرتب کنید
۴۷	۴-۴ بسته بندی (مجموعه) داده ها را بررسی کنید
۵۰	۴-۵ به ابتدا و انتهای داده ها نگاه کنید
۵۱	۴-۶ تعداد یا «n» ها را بررسی کنید
۵۷	۴-۷ با حداقل یک منبع داده خارجی اعتبارسنجی کنید
۵۹	۴-۸ نمودار رسم کنید
۶۱	۴-۹ ابتدا راه حل آسان را امتحان کنید
۶۵	۴-۱۰ پیگیری کنید و ادامه دهید
۶۷	۵- بهره گیری از مدل ها برای کاوش داده ها
۶۹	۵-۱ مدل های پیش بینی
۷۱	۵-۲ مقایسه پیش بینی مدل با واقعیت
۷۴	۵-۳ واکنش به داده ها: پالایش انتظارات
۷۶	۵-۴ آزمایش روابط خطی
۸۱	۵-۵ چه زمانی توقف کنیم؟
۸۵	۵-۶ خلاصه

۸۷	۶- استنباط.....
۸۷	۶-۱ جمعیت را شناسایی کنید.....
۸۸	۶-۲ فرایند نمونه‌گیری را شرح دهید.....
۸۸	۶-۳ مدلی برای جمعیت توصیف کنید.....
۸۹	۶-۴ یک مثال ساده.....
۹۲	۶-۵ عوامل مؤثر بر کیفیت استنباط.....
۹۳	۶-۶ مثال: کاربرد موسیقی ایل.....
۹۷	۶-۷ جمعیت به اشکال مختلفی وجود دارد.....
۱۰۱	۷- مدل‌سازی رسمی.....
۱۰۱	۷-۱ اهداف مدل‌سازی رسمی چیست؟.....
۱۰۲	۷-۲ چارچوب کلی.....
۱۰۴	۷-۳ تجزیه و تحلیل ارتباطی.....
۱۱۳	۷-۴ تجزیه و تحلیل پیش‌بینی.....
۱۲۰	۷-۵ خلاصه.....
۱۲۱	۸- استنباط در مقابل پیش‌بینی: پیامدهای مربوط به استراتژی مدل‌سازی.....
۱۲۲	۸-۱ آلودگی هوا و مرگ و میر در شهر نیویورک.....
۱۲۴	۸-۲ استناد به ارتباطات.....
۱۲۹	۸-۳ پیش‌بینی خروجی.....
۱۳۰	۸-۴ خلاصه.....
۱۳۳	۹- تفسیر نتایج.....
۱۳۴	۹-۱ مطالعه موردی: مصرف نوشابه غیررژیمی و شاخص توده بدنی (BMI).....
۱۵۳	۱۰- گفتگو(ارتباطات).....
۱۵۴	۱۰-۱ ارتباطات متداول.....
۱۵۵	۱۰-۲ مخاطب.....
۱۵۷	۱۰-۳ محتوا.....
۱۵۹	۱۰-۴ سبک.....
۱۶۰	۱۰-۵ نگرش.....
۱۶۱	۱۰-۶ اندیشه‌های پایانی.....
۱۶۳	۱۱- درباره نویسندگان کتاب.....