

داده‌کاوی با پایتون به همراه تحلیل شبکه‌های اجتماعی

مؤلفان:

دکتر بابک تیمورپور

(عضو هیأت علمی دانشگاه تربیت مدرس)

وحید یادگاری

محمد فاتحی پیکانی



علوم رایانه

سرشناسه	: تیمورپور، بابک، ۱۳۵۰ -
عنوان و نام پدیدآور	: داده‌کاوی با پایتون به همراه تحلیل شبکه‌های اجتماعی/مولفان بابک تیمورپور، وحید یادگاری، محمد فاتحی پیکانی.
مشخصات نشر	: بابل: علوم رایانه، ۱۳۹۹.
مشخصات ظاهری	: ۱۷۶ ص.: جدول، نمودار.
شابک	: ۹۷۸-۶۰۰-۲۰۵-۱۶۹-۱-۱
وضعیت فهرست نویسی	: فیبا
یادداشت	: کتابنامه.
موضوع	: پایتون (زبان برنامه‌نویسی کامپیوتر) : Python (Computer program language)
موضوع	: داده‌کاوی : Data mining
موضوع	: شبکه‌های اجتماعی پیوسته : Online social networks
شناسه افزوده	: یادگاری، وحید، ۱۳۵۸ -
شناسه افزوده	: فاتحی پیکانی، محمد، ۱۳۶۶ -
رده بندی کنگره	: QA ۷۶/۷۳
رده بندی دیویی	: ۰۵/۱۳۳
شماره کتابشناسی ملی	: ۷۵۲۸۱۴۸
وضعیت رکورد	: فیبا

این اثر، مشمول قانون حمایت مولفان و مصنفان و هم‌نقدان مصوب ۱۳۴۸ است، هرکس تمام یا قسمتی از این اثر را بدون اجازه‌ی مولف (ناشر) نشر یا پخش یا عرضه کند مورد پیگرد قانونی قرار خواهد گرفت.

تلفن : ۰۱۱ - ۳۲۳۶۰۷۷۲

www.olomrayaneh.net
بابل، صندوق پستی ۸۹۱-۴۷۱۳۵



علوم رایانه

داده‌کاوی با پایتون به همراه تحلیل شبکه‌های اجتماعی
تألیف: دکتر بابک تیمورپور - وحید یادگاری - محمد فاتحی پیکانی

چاپ اول

زمستان ۱۳۹۹

شمارگان: ۵۰۰ نسخه

قیمت: ۵۵۰۰۰ تومان

چاپ و صحافی: چاپ دیجیتال میلاد بابل

شابک: ۹۷۸-۶۰۰-۲۰۵-۱۶۹-۱-۱

نشانی: بابل، خیابان شریعتی، مجتمع میلاد، واحد ۱۷

حروفچینی و صفحه‌آرایی: علوم رایانه

تهران، خیابان انقلاب، خیابان منیری جاوید، نبش وحید نظری، شماره ۱۴۲ - تلفکس: ۰۲۲۰-۶۶۴۰۰۱۴۴-۶۶۴۰۰۱۴۴

فهرست مطالب

فصل اول : مقدمه‌ای بر داده‌کاوی

- ۱-۱. داده‌کاوی و کشف دانش در پایگاه داده‌ها ۹
- ۲-۱. فرآیند کشف دانش ۱۰
- ۳-۱. عملکردهای داده‌کاوی ۱۳
- ۴-۱. کاربردهای داده‌کاوی ۱۳
- ۵-۱. انواع حوزه‌های اطلاعات ۱۴
- ۶-۱. انواع داده ۱۶
- ۷-۱. داده، اطلاعات و دانش ۲۰
- ۸-۱. پایگاه داده ۲۵

فصل دوم : مقدمه‌ای بر پایتون

- ۱-۲. مقدمات و توابع ابتدایی ۲۷
- ۲-۲. نصب و اجرای پایتون ۲۸
- ۳-۲. اجرای پایتون در محیط توسعه‌ی یکپارچه ۲۹
- ۴-۲. توزیع آنکوندا ۳۱
- ۵-۲. عملگرها در پایتون ۳۳
- ۶-۲. انواع داده در پایتون ۳۵
- ۷-۲. تابع در پایتون ۴۰
- ۸-۲. حلقه‌ی for ۴۱
- ۹-۲. تابع range() ۴۲
- ۱۰-۲. تابع zip() ۴۳
- ۱۱-۲. تابع بی‌نام ۴۳
- ۱۲-۲. بسته‌های پایتون ۴۳
- ۱۳-۲. آمار توصیفی در پایتون با کتابخانه‌ی Numpy ۴۴
- ۱۴-۲. ساختار قاب داده و کتابخانه‌ی Pandas ۴۵
- ۱۵-۲. مدیریت مقادیر گمشده ۵۱
- ۱۶-۲. کار با داده و فایل‌های اکسل در پایتون ۵۳
- ۱۷-۲. مصورسازی داده‌ها ۵۳
- ۱۸-۲. نمودار میله‌ای ۵۳
- ۱۹-۲. کتابخانه‌ی Scikit-learn ۵۵

فصل سوم: تحلیل اکتشافی داده‌ها در پایتون

- ۱-۳. آماده‌سازی داده‌ها برای داده‌کاوی ۵۶
- ۲-۳. پیش‌پردازش داده‌ها ۵۶
- ۳-۳. تحلیل اکتشافی داده‌ها در پایتون ۶۳
- ۴-۳. کاهش بُعد ۶۶
- ۵-۳. تشخیص داده‌های پرت با استفاده از خوشه‌بندی ۶۸

فصل چهارم: خوشه‌بندی

- ۱-۴. خوشه‌بندی ۶۹
- ۲-۴. تفاوت دسته‌بندی و خوشه‌بندی ۶۹
- ۳-۴. اعتبارسنجی خوشه‌ها ۷۰
- ۴-۴. روش‌های خوشه‌بندی ۷۱
- ۵-۴. شاخص‌های اعتبارسنجی ۷۶
- ۶-۴. پیاده‌سازی الگوریتم‌های خوشه‌بندی ۷۶

فصل پنجم: دسته‌بندی و پیش‌بینی

- ۱-۵. دسته‌بندی ۸۹
- ۲-۵. راه متداول ساخت مدل دسته‌بندی ۹۰
- ۳-۵. تفاوت دسته‌بندی و خوشه‌بندی ۹۰
- ۴-۵. ارزیابی مدل‌ها ۱۰۵
- ۵-۵. دسته‌بندی در پایتون ۱۰۷
- ۶-۵. پیش‌بینی در پایتون ۱۱۳
- ۷-۵. قوانین انجمنی ۱۱۷

فصل ششم: تحلیل شبکه‌های اجتماعی

- ۱-۶. شبکه‌های اجتماعی ۱۲۳
- ۲-۶. انواع مرکزیت و شاخص‌های اصلی در تحلیل شبکه ۱۲۵
- ۳-۶. پیش‌بینی شکل‌گیری ارتباط ۱۲۹
- ۴-۶. اجتماع‌یابی در شبکه ۱۳۰
- ۵-۶. هم‌ارزی ساختاری ۱۳۱
- ۶-۶. ضریب خوشه‌بندی ۱۳۳
- ۷-۶. انتقال‌پذیری ۱۳۴
- ۸-۶. الگوریتم PageRank ۱۳۵
- ۹-۶. تحلیل شبکه‌ای در پایتون ۱۳۶
- ۱۰-۶. محاسبه‌ی شاخص‌های کلان شبکه ۱۳۹

پیش‌گفتار

داده‌کاوی یکی از مفاهیم نوین میان رشته‌ای است که داده‌های حوزه‌های مختلفی همچون ریاضیات، آمار، فناوری اطلاعات، مدیریت و سایر زمینه‌های مرتبط را با هم ترکیب می‌کند تا اطلاعات و دانش مفید و موثر پنهان در حجم بزرگی از داده‌ها را استخراج نماید. امروزه با گسترش فزاینده‌ی فناوری اطلاعات، تقریباً همه‌ی سازمان‌ها حجم عظیمی از داده‌ها را در پایگاه داده‌ی خود ذخیره می‌کنند. داده‌کاوی این پایگاه‌ها را متعاقب کشف و استخراج دانش با استفاده از روش‌های یادگیری ماشین تحلیل می‌کند.

زبان پایتون با وجود کتابخانه‌های گسترده و به‌روز خود در حوزه‌ی علوم داده، یکی از زبان‌های برنامه‌نویسی قدرتمند و کارآمد در داده‌کاوی است. پایتون به دلیل سادگی و همه‌منظوره بودن و ایجاد برنامه‌های کاربردی و تحلیل داده مورد توجه همگان قرار گرفته است. همچنین داشتن کتابخانه‌های متعدد و دسترسی آسان به آن موجب گرایش بسیاری از متخصصان داده و برنامه‌نویسان شده است.

علاقه‌مندان می‌توانند برنامه‌های کتاب را از سایت انتشارات علوم رایانه با آدرس www.olomrayaneh.net به طور رایگان دریافت کنند. برای این کار، در صفحه‌ی اول سایت به لینک "کدهای رایگان بروید" و این کتاب را انتخاب کنید.

امید است این اثر درجه‌ای نو به فهم مسائل مرتبط به تجزیه و تحلیل داده‌ها و اطلاعات به روی مخاطبان ارجمند بگشاید و برای ارتقای سطح دانش کشور مفید واقع شود.