



به نام خدا



۳۳۰۷۸۰۵

۱۴۰۰۴۲۹

پروژه های کاربردی علم داده در پایتون

www.ketab.ir

مؤلف:

جاوید مولاپور



هرگونه چاپ و تکثیر از محتویات این کتاب بدون اجازه کتبی ناشر ممنوع است. متخلفان به موجب قانون حمایت حقوق مؤلفان، مصنفان و هنرمندان تحت پیگرد قانونی قرار می‌گیرند.

◀ عنوان کتاب: پروژه های کاربردی علم داده در پایتون

◀ مولف : جاوید مولاپور

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ ویراستار: نرگس مهرید

◀ صفحه آرای: فرناز عبدالهی

◀ طراح جلد: داریوش فرسای

◀ نوبت چاپ: دوم

◀ تاریخ نشر: ۱۴۰۰

◀ چاپ و صحافی: درج عقیق

◀ تیراژ: ۵۰ جلد

◀ قیمت: ۹۳۰۰۰۰ ریال

◀ شابک: ۹۷۸-۶۲۲-۲۱۸-۴۳۹-۱

◀ نشانی واحد فروش: تهران، میدان انقلاب،

خ کارگر جنوبی، روبروی پاساژ مهستان،

پلاک ۱۲۵۱-تلفن: ۰۴۶-۶۶۴۱۰۰۴۶-۲۲۰۸۵۱۱۱

◀ فروشگاههای اینترنتی دیباگران تهران :

WWW.MFTBOOK.IR

www.dibagaran-tehran.com

◀ نشانی تلگرام: @mftbook

◀ نشانی اینستاگرام دیبا [dibagaran_publishing](https://www.instagram.com/dibagaran_publishing)

هر کتاب دیباگران، یک فرصت جدید شغلی و علمی.

هرگوشی همراه، یک فروشگاه کتاب دیباگران تهران.

از طریق سایتهای دیباگران، در هر جای ایران به کتابهای ما دسترسی دارید.

سرشناسه: مولاپور اندیلی، جاوید، ۱۳۶۶-
عنوان و نام پدیدآور: پروژه های کاربردی علم داده در
پایتون / مولف: جاوید مولاپور؛ ویراستار: نرگس مهرید.
مشخصات نشر: تهران: دیباگران تهران، ۱۴۰۰
مشخصات ظاهری: ۲۴۰ ص: مصور.
شابک: ۹۷۸-۶۲۲-۲۱۸-۴۳۹-۱
ویراست: فهرست نویسی: فیبا
موضوع: پایتون زبان برنامه نویسی کامپیوتر
موضوع: Python (computer program language)
موضوع: زبان شناسی کامپیوتری
موضوع: computational linguistics
موضوع: داده کاوی data mining
رده بندی کنگره: QA ۷۶/۷۳
رده بندی دیویی: ۰۰۵/۱۳۳
شماره کتابشناسی ملی: ۷۶۴۶۲۱۹

فهرست مطالب

۱۲	فصل اول / آشنایی با علم داده
۱۳	مقدمه
۱۳	چرا داده‌ها مهم هستند؟
۱۳	علم داده چیست؟
۱۴	مزایای علم داده
۱۷	انتخاب محیط توسعه
۲۱	مقدمه‌ای بر سری‌ها و DATAFRAME ها در PANDAS
۲۱	تمرین کار با مجموعه داده‌ها
۲۲	ایمپورت کردن کتابخانه‌ها و مجموعه داده
۲۲	اکتشاف داده سریع
۲۳	تحلیل توزیع
۲۶	تحلیل متغیرهای دسته‌ای
۲۸	پیش‌پرداز داده‌ها (DATA MUNGING) در پایتون با استفاده از PANDAS
۲۸	بررسی مقادیر ناموجود در مجموعه داده
۳۱	ساخت یک مدل پیش‌بین در پایتون
۳۲	رگرسیون لجستیک
۳۳	درخت تصمیم
۳۴	جنگل تصادفی
۳۵	سری زمانی در علم داده
۳۶	تابع خودهمبستگی (AUTOCORRELATION FUNCTION)
۳۶	تغییرات فصلی (SEASONALITY)
۳۷	ایستایی (STATIONARY)
۳۷	بررسی ایستایی سری زمانی
۳۸	مدل‌سازی سری زمانی
۳۸	مدل میانگین متحرک (MOVING AVERAGE)
۳۹	هموارسازی نمایی
۴۰	هموارسازی نمایی مضاعف (DOUBLE EXPONENTIAL SMOOTHING)
۴۱	هموارسازی نمایی سه‌تایی (TRIPLE EXPONENTIAL SMOOTHING)
۴۳	پروژه پیش‌بینی سری زمانی
۴۳	مبنای کارایی پیش‌بینی

۴۵..... پروژه پیش‌بینی سری زمانی در پایتون و الگوریتم مانا

۴۹..... فصل دوم / آشنایی با PANDAS

۵۰..... PANDAS

۵۰..... مزایای استفاده از پانداس

۵۰..... SERIES در پانداس

۵۲..... وارد کردن داده‌های CSV

۵۴..... وارد کردن داده‌های اکسل

۵۵..... دیتافریم DATAFRAME

۵۶..... منظور از DATAFRAME چیست؟

۵۶..... چگونه یک قاب داده PANDAS ایجاد کنیم؟

۵۸..... مشاهده داده‌های یک قاب

۶۰..... بارگزاری ستون‌ها در قاب داده

۶۰..... حذف ردیف‌ها و ستون‌ها از قاب داده

۶۱..... ایجاد یک قاب داده PANDAS از یک لیست

۶۲..... عملیات ریاضی روی قاب داده

۶۳..... پیش‌پردازش داده‌ها

۶۴..... GROUPBY در پانداس

۶۵..... کاربردهای پایه‌ای GROUPBY در پایتون

۶۶..... استفاده از تابع سفارشی در GROUPBY پانداس

۶۷..... عملیات روی گروه‌های پانداس

۶۷..... تکرار و انتخاب گروه‌ها

۶۸..... متد GET_GROUP در پانداس

۶۹..... متد VALUE_COUNTS در پایتون

۶۹..... آمار توصیفی

۷۰..... پاکسازی داده‌ها با استفاده از NUMPY و PANDAS

۷۱..... حذف ستون‌ها در یک DATAFRAME

۷۳..... تغییر INDEX دیتافریم

۷۴..... مرتب‌سازی فیلدهای داده

۷۵..... ترکیب متدهای STR با NUMPY برای پاکسازی ستون‌ها

۷۸..... پاکسازی کل مجموعه داده با استفاده از تابع APPLYMAP

۸۱..... تغییر نام ستون‌ها و گذر از سطرها

۸۴..... فصل سوم / آشنایی با TENSORFLOW

۸۵..... TENSORFLOW چیست؟

۸۵..... کاربرد تانسورفلو

۸۷..... ویژگی‌های TENSORFLOW