



به نام خدا



پروژه های کاربردی علم داده در پایتون

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 مولا پور

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ ویراستار: نرگس مهربد

◀ صفحه آرای: فرنوش عبدالمهی

◀ طراح جلد: داریوش فرسای

◀ نوبت چاپ: اول

◀ تاریخ نشر: ۱۴۰۰

◀ چاپ و صحافی: صدف

◀ تیراژ: ۱۰۰ جلد

◀ قیمت: ۸۳۰۰۰۰ 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.dibagarantehran.com

www.dibbook.ir

نشانی اینستاگرام دیبا [@dibagaran_publishing](https://www.instagram.com/dibagaran_publishing) نشانی تلگرام: [@mftbook](https://www.telegram.com)

هر کتاب دیباگران، یک فرصت جدید شغلی و علمی.

هر گوشی همراه، یک فروشگاه کتاب دیباگران تهران.

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

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..... بارگذاری 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 PANDAS و NUMPY

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 تنسورفلو

۸۷..... ویژگی‌های TENSORFLOW

۸۹	معماری تنسورفلو
۸۹	TENSORFLOW SERVABLE
۹۲	مبانی تنسورفلو
۹۲	نصب TENSORFLOW
۹۴	ساخت تنسور N بعدی
۹۶	شکل تنسور
۹۷	انواع داده
۹۸	ایجاد اپراتور (آشنایی با اپراتورهای ریاضی)
۱۰۳	گراف
۱۰۴	مراحل ایجاد PIPELINE در تنسورفلو
۱۰۹	گراف محاسباتی
۱۱۰	اجرای موازی در گرافهای محاسباتی
۱۱۰	اجرای توزیع شده یا غیرمتمرکز
۱۱۱	زیرگرافهای محاسباتی
۱۱۱	لزوم فشرده سازی داده ها
۱۱۲	انواع تنسور در فریم ورک تنسورفلو
۱۱۳	ایجاد یک تنسور N بعدی
۱۱۶	متغیرها در تنسورفلو
۱۱۷	PLACEHOLDERها در تنسورفلو
۱۱۸	مفهوم SESSION در تنسورفلو
۱۲۰	اجرا روی پردازنده یا کارت گرافیکی
۱۲۱	آموزش مدل رگرسیون خطی
۱۲۴	نحوه کارکرد الگوریتم رگرسیون خطی
۱۲۵	چگونه یک رگرسیون خطی را با TENSORFLOW آموزش می دهیم؟
۱۲۷	رگرسیون خطی به کمک PANDAS
۱۳۲	راه حل مبتنی بر TENSORFLOW
۱۳۹	فصل چهارم / آموزش کتابخانه SCIPY
۱۴۰	SUB-PACKAGE های SCIPY چیست؟
۱۴۰	ساختار داده های SCIPY چیست؟
۱۴۱	توابع پایه ای در SCIPY چیست؟
۱۴۱	ماتریس صفر
۱۴۲	ماتریس یک
۱۴۲	تابع ARANGE
۱۴۲	تعیین نوع مقادیر
۱۴۲	ماتریس

۱۴۳	خوشه‌بندی یا CLUSTERING در SCIPY
۱۴۳	اجرای K-MEANS با SCIPY
۱۴۴	محاسبه KMEANS با سه خوشه
۱۴۵	مقادیر ثابت در SCIPY چیست؟
۱۴۵	لیست مقادیر ثابت موجود
۱۴۶	تبدیل سریع فوریه
۱۴۶	تبدیل فوریه گسسته یک‌بعدی
۱۴۸	شکل سیگنال سینوسی SCIPY - چیست؟
۱۴۹	کاربرد کتابخانه SCIPY
۱۵۰	جبر خطی در SCIPY
۱۵۰	تفاوت جبر خطی در NUMPY و SCIPY چیست؟
۱۵۰	معادلات خطی
۱۵۱	محاسبه دترمینان
۱۵۳	پردازش تصویر با SCIPY
۱۵۳	باز کردن و نوشتن در فایل‌های تصویری
۱۵۶	فیلترگذاری در تصاویر
۱۵۶	محو و تارشدن تصویر
۱۵۷	تشخیص لبه‌های تصویر EDGE DETECTION
۱۵۹	بهینه‌سازی با SCIPY
۱۵۹	الگوریتم مینیمم‌سازی NELDER – MEAD
۱۶۰	محاسبه حداقل مربعات
۱۶۰	جمع‌بندی

فصل پنجم / کار با کتابخانه OPENPYXL

۱۶۲	مقدمه
۱۶۲	نصب OPENPYXL
۱۶۳	وارد کردن اطلاعات به فایل موجود
۱۶۴	خواندن اطلاعات از فایل
۱۶۵	خواندن اطلاعات چند سلول
۱۶۶	تابع ITER_ROW()
۱۶۷	متد ITER_COL()
۱۶۷	کار با چند SHEET
۱۶۸	فیلتر و مرتب‌سازی داده‌ها
۱۶۹	ادغام کردن سلول‌ها
۱۷۰	استفاده از فرمول‌های اکسل
۱۷۱	اضافه کردن نمودار به فایل اکسل

۱۷۳	افزودن تصویر
فصل ششم / پروژه تشخیص رنگ از روی تصویر با OPENCV و PANDAS ۱۷۴	
۱۷۵	مقدمه
۱۷۵	پیش‌نیازها
۱۷۷	OPENCV
۱۷۷	کاربردهای پردازش تصویر
۱۷۸	برنامه‌نویسی برای پردازش تصویر
۱۷۸	تاریخچه
۱۷۸	چگونه کامپیوتر تصویر را تشخیص می‌دهد؟
۱۸۰	چرا OPENCV برای COMPUTER VISION استفاده می‌شود؟
۱۸۰	نصب کتابخانه‌های OPENCV, PANDAS, NUMPY
۱۸۱	ساخت تابع DRAW_FUNCTION
۱۸۴	سورس کامل برنامه
فصل هفتم / تشخیص جنسیت و تخمین سن از روی تصویر با استفاده از CNN, OPENCV ... ۱۸۶	
۱۸۷	مقدمه
۱۸۸	لایه‌های همگرا
۱۸۹	لایه‌های ادغام
۱۸۹	شبکه‌های عصبی متأخر
۱۹۰	تاریخچه
۱۹۱	بینایی ماشین
۱۹۲	وظایف اصلی در بینایی رایانه‌ای
۱۹۳	استخراج ویژگی
۱۹۴	بینایی و تفسیر تصاویر در انسان‌ها
۱۹۶	موارد استفاده از تکنولوژی COMPUTER VISION
۱۹۷	سرویس‌های شناختی IBM
۱۹۷	کاربردهای نظامی
۱۹۸	درباره پروژه
۱۹۸	معماری CNN
۱۹۹	مراحل انجام پروژه برای تشخیص جنسیت و سن
۲۰۳	سورس کامل برنامه

۲۰۷	درباره پروژه
۲۰۷	مقدمه
۲۰۸	شبکه‌های عصبی
۲۰۹	تعریف
۲۰۹	کارکرد
۲۱۰	انواع شبکه‌های عصبی مصنوعی
۲۱۰	پرسترون چندلایه یا MLP
۲۱۱	شبکه‌های عصبی شعاعی یا RBF
۲۱۲	ماشین‌های بردار پشتیبان یا SVM
۲۱۲	نگاشت‌های خودسازمان‌ده یا SOM
۲۱۳	یادگیرنده رقمی‌ساز بردار یا LVQ
۲۱۳	شبکه عصبی هاپفیلد یا HOPFIELD
۲۱۴	واحد پردازش تنسور (TPU)
۲۱۵	پیکسل‌ویژوال کور (PVC)
۲۱۵	کاربردها
۲۱۶	کتابخانه PILLOW
۲۱۶	KERAS
۲۱۷	پروژه تشخیص اعداد دستنویس
۲۱۷	مرحله اول: پیش‌نیازها
۲۱۷	مرحله دوم: وارد کردن کتابخانه‌ها و بارگیری DATASET
۲۱۸	مرحله سوم: پردازش داده‌ها
۲۱۸	مرحله چهارم: ساخت مدل
۲۱۹	مرحله پنجم: آموزش مدل
۲۱۹	مرحله ششم: ارزیابی مدل
۲۱۹	مرحله هفتم: ساخت رابط کاربری برنامه
۲۲۱	اجرای برنامه
۲۲۲	فصل نهم / کار با صدا در علم داده
۲۲۳	مقدمه
۲۲۳	پروژه یادگیری عمیق
۲۲۵	تعریف مدل کرس
۲۲۶	کامپایل کردن مدل KERAS
۲۲۶	برازش مدل KERAS
۲۲۷	ارزیابی مدل KERAS
۲۲۷	یکپارچه‌سازی کلیه موارد
۲۲۹	انجام پیش‌بینی

۲۳۱	دسته بندی صدا با DEEP LEARNING
۲۳۱	مجموعه داده
۲۳۲	اکتشاف داده
۲۳۴	پیش پردازش داده‌ها
۲۳۵	استخراج ویژگی‌ها
۲۳۶	تبدیل داده‌ها و برچسب‌ها و سپس تقسیم مجموعه داده
۲۳۷	ساخت مدل
۲۳۹	نتایج
۲۳۹	مشاهدات

خط‌مشی انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب‌هایی با کیفیت عالی است که تواند
خواسته‌های به روز جامعه فرهنگی و علمی کشور را تا حد امکان پوشش دهد.
هر کتاب دیباگران تهران، یک فرصت جدید شغلی و علمی

حمد و سپاس ایزد منان را که با الطاف بی‌کران خود این توفیق را به ما ارزانی داشت تا بتوانیم در راه ارتقای دانش عمومی و فرهنگی این مرز و بوم در زمینه چاپ و نشر کتب علمی و آموزشی گام‌هایی هرچند کوچک برداشته و در انجام رسالتی که بر عهده داریم، مؤثر واقع شویم.

گسترده‌گی علوم و سرعت توسعه روزافزون آن، شرایطی را به وجود آورده که هر روز شاهد تحولات اساسی چشمگیری در سطح جهان هستیم. این گسترش و توسعه، نیاز به منابع مختلف از جمله کتاب را به عنوان قدیمی‌ترین و راحت‌ترین راه دستیابی به اطلاعات و اطلاع‌رسانی، بیش از پیش برجسته نموده است.

در این راستا، واحد انتشارات مؤسسه فرهنگی هنری دیباگران تهران با همکاری اساتید، مؤلفان، مترجمان، متخصصان، پژوهشگران و محققان در زمینه‌های گوناگون و مورد نیاز جامعه تلاش نموده برای رفع کمبودها و نیازهای موجود، منابعی پُر بار، معتبر و با کیفیت مناسب در اختیار علاقمندان قرار دهد.

کتابی که در دست دارید با همت "جناب آقای جاوید مولاپور" و تلاش جمعی از همکاران انتشارات میسر گشته که شایسته است از یکایک این گرامیان تشکر و قدردانی کنیم.

با نظرات خود مشوق و راهنمای ما باشید

با ارائه نظرات و پیشنهادات و خواسته‌های خود، به ما کمک کنید تا بهتر و دقیق‌تر در جهت رفع نیازهای علمی و آموزشی کشورمان قدم برداریم. برای رساندن پیام‌هایتان به ما از انواع رسانه‌های دیباگران تهران شامل سایتهای فروشگاهی و صفحه اینستاگرام و شماره‌های تماس که در صفحه شناسنامه کتاب آمده استفاده نمایید.

مدیر انتشارات

مؤسسه فرهنگی هنری دیباگران تهران
bookmarket@mft.info

❖ مقدمه مؤلف

هنگامی که سازمان‌ها با حجم انبوهی از داده‌ها مواجه هستند، نیاز است تا بتوانند از این داده‌ها و اطلاعات بزرگ استفاده کنند و بر این اساس استراتژی کاری اعم از مالی، بازاریابی، سرمایه گذاری و ... خود را بهبود بخشند، در این مرحله است که علم داده پا به عرصه می‌گذارد. علم داده یا دیتا ساینس یکی از مباحث روز دنیا می‌باشد، که با استفاده از کامپیوتر و فناوری اطلاعات شکل گرفته‌است. این حوزه اساساً متکی به علم کامپیوتر می‌باشد. جذابیت علم داده به حدی است که امروزه در بیش‌تر دانشگاه‌های دنیا دوره‌های تخصصی برای تدریس آن در نظر گرفته شده‌است. ضمن اینکه پژوهش‌های زیادی در این زمینه رو به افزایش می‌باشد. از این رو سعی شده‌است در این کتاب به صورت کاملاً کاربردی به مباحث پرداخته و پروژه‌های کاربردی نیز پیاده‌سازی گردد.

این کتاب مشتمل بر موضوعات کاربردی علم داده به صورت پروژه‌محور و ساخت چندین پروژه از جمله کار بر روی داده‌ها، سری زمانی، پاکسازی داده‌ها، اکتشاف، پیش‌بینی، شبکه‌های عصبی، یادگیری عمیق، کار بر روی تصاویر و صدا می‌باشد. جهت کار با پروژه‌های این کتاب آشنایی اولیه با زبان برنامه‌نویسی پایتون می‌تواند جهت تسریع در روال یادگیری مفید باشد. هم‌چنین از زبان پایتون نسخه ۳.۷ جهت پیاده‌سازی پروژه‌های این کتاب استفاده شده‌است.

در پایان از خوانندگان محترم تقاضا می‌شود نظرات، پیشنهادات و یا در صورت مشاهده هر گونه اشکال در کتاب اینجانب را آگاه سازند.

جاوید مولاپور

molapour.javid@gmail.com