



به نام خدا

کتابخانه های پایتون

برای هوش مصنوعی و یادگیری ماشین

مؤلف:

نفسه رضانی پور



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: **کتابخانه های پایتون**

برای هوش مصنوعی و یادگیری ماشین

◀ مولف: نفیسه رضانی پور

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ ویراستار: نرگس مهربد

◀ صفحه آرای: فریوش عبدالهی

◀ طراح جلد: داریوش فرسای

◀ نوبت چاپ: اول

◀ تاریخ نشر: ۱۴۰۲

◀ چاپ و صحافی: صدف

◀ تیراژ: ۱۰۰ جلد

◀ قیمت: ۱۹۸۰۰۰۰ ریال

◀ شابک: ۹۷۸-۶۲۲-۲۱۸-۶۶۴-۷

◀ نشانی واحد فروش: تهران، خیابان انقلاب، خیابان دانشگاه

◀ تقاطع شهدای ژاندارمری - پلاک ۱۵۸ ساختمان دانشگاه -

◀ طبقه دوم - واحد ۴ تلفن ها: ۶۶۹۶۵۷۴۹-۲۲۰۸۵۱۱۱

◀ فروشگاههای اینترنتی دیباگران تهران :

WWW.MFTBOOK.IR

www.dibagaran.com

سرشناسه: رضانی پور، نفیسه، ۱۳۶۱-

عنوان و نام پدیدآور: کتابخانه های پایتون برای هوش

مصنوعی و یادگیری ماشین / مولف: نفیسه رضانی پور؛

ویراستار: نرگس مهربد.

مشخصات نشر: تهران: دیباگران تهران: ۱۴۰۱

مشخصات ظاهری: ۱۹۸ ص: مصور، جدول، نمودار.

شابک: ۹۷۸-۶۲۲-۲۱۸-۶۶۴-۷

وضعیت فهرست نویسی: فیبا یادداشت: کتابنامه: ۱۹۷-۱۹۸

موضوع: پایتون (زبان برنامه نویسی کامپیوتر)

موضوع: Python (computer program language)

موضوع: هوش مصنوعی - کاربرد کتابخانه ای

موضوع: Artificial intelligence-library

applications

رده بندی کنگره: ۷۶/۷۳ QA

رده بندی دیویی: ۰۰۵/۱۳/۳

شماره کتابشناسی ملی: ۹۲۳۳۶۶۲

نشانی تلگرام: @mftbook نشانی اینستاگرام دیبا dibagaran_publishing

هر کتاب دیباگران، یک فرصت جدید علمی و شغلی.

هر گوشی همراه، یک فروشگاه کتاب دیباگران تهران.

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

فهرست مطالب

فصل اول / یادگیری ماشین چیست؟ و معرفی ابزارهای مورد نیاز ۱۲

۱۳..... هوش مصنوعی چیست؟

۱۳..... یادگیری ماشین چیست؟

۱۴..... گوگل کولب

۱۴..... ساخت اکانت گوگل کولب

۱۵..... اجرای کد

۱۶..... اجرای تمامی سلول‌ها

۱۶..... افزودن سلول جدید

۱۶..... حذف سلول

۱۶..... افزودن متن در کولب

۱۷..... ذخیره نوتبوک

۱۸..... اجرا بر روی GPU

۱۸..... بارگذاری یک فایل در گوگل درایو

۱۹..... تغییر تم نوتبوک

فصل دوم / آشنایی با کتابخانه NUMPY ۲۰

۲۱..... NUMPY

۲۱..... ساخت آرایه

۲۲..... ساخت آرایه با ابعاد و مقادیر دلخواه با دستور *np.full()*

۲۲..... ساخت آرایه با دستور *arange*

۲۳..... ساخت آرایه با دستور *linspace*

۲۳..... شکل آرایه‌ها با دستور *shape*

۲۴..... ابعاد آرایه با دستور *ndim*

۲۴..... تعداد عناصر آرایه با دستور *size*

۲۴..... نوع داده‌ای عناصر آرایه با دستور *dtype*

۲۴..... تغییر شکل آرایه‌ها

۲۵.....	ستونی
۲۵.....	سطری
۲۵.....	ایجاد بردار یک بعدی
۲۶.....	مرتب‌سازی عناصر آرایه
۲۶.....	پیدا کردن عناصر یکتا در آرایه
۲۷.....	محاسبه واریانس، انحراف معیار و میانگین و میانه
۲۷.....	میانگین
۲۷.....	واریانس
۲۷.....	انحراف معیار
۲۷.....	میانه
۲۸.....	ترانهاده یک آرایه
۲۸.....	جمع عناصر آرایه
۲۸.....	پیدا کردن بیشترین و کمترین مقدار آرایه‌ها
۲۹.....	ضرب ۲ آرایه به صورت نظیر به نظیر
۲۹.....	جمع ۲ آرایه به صورت نظیر به نظیر
۲۹.....	تفریق ۲ آرایه به صورت نظیر به نظیر
۳۰.....	تقسیم ۲ آرایه به صورت نظیر به نظیر
۳۰.....	ساخت پشته با استفاده از آرایه‌ها
۳۱.....	تولید اعداد تصادفی با استفاده از کتابخانه NUMPY
۳۲.....	تغییر ترتیب اعداد یک آرایه
۳۳.....	تولید یک بازه از اعداد بدون نظم مشخص
۳۳.....	ساخت ماتریس
۳۴.....	تولید ماتریس همانی
۳۵.....	عملیات روی ماتریس‌ها
۳۵.....	ترانهاده ماتریس
۳۵.....	ضرب ماتریسی
۳۶.....	ضرب نظیر به نظیر درایه‌های ماتریس
۳۶.....	جمع نظیر به نظیر درایه‌های ماتریس
۳۶.....	تفریق نظیر به نظیر درایه‌های ماتریس
۳۶.....	تقسیم نظیر به نظیر درایه‌های ماتریس
۳۷.....	تفاوت <i>matrix</i> و <i>ndarray</i>

فصل سوم / کار با داده‌ها و آشنایی با کتابخانه‌های SCIKIT-LEARN و PANDAS..... ۳۸

۳۹.....	هوش مصنوعی
۳۹.....	یادگیری ماشین
۳۹.....	علم داده
۴۰.....	آشنایی با کتابخانه (PANDAS (PYTHON DATA ANALYSIS LIBRARY)
۴۰.....	آشنایی با کتابخانه SCIKIT-LEARN
۴۱.....	ساختار استاندارد داده
۴۱.....	انواع داده
۴۲.....	داده‌های دسته‌بندی
۴۲.....	روش‌های بارگذاری داده‌ها
۴۳.....	بارگذاری داده از کتابخانه <i>scikit-learn</i>
۴۴.....	تولید داده شبیه‌سازی شده با استفاده از کتابخانه <i>scikit-learn</i>
۴۵.....	خواندن داده از یک فایل <i>csv</i>
۴۸.....	خواندن داده از یک آدرس <i>url</i>
۴۹.....	خواندن داده از درایو سیستم
۴۹.....	خواندن داده از یک فایل <i>json</i>
۵۱.....	توصیف داده‌ها
۵۲.....	نمایش چند سطر از داده‌ها
۵۲.....	تبدیل داده به <i>Data Frame</i>
۵۲.....	ساخت دیتافریم با استفاده از دیکشنری
۵۳.....	ساخت دیتافریم با استفاده از آرایه
۵۳.....	نحوه دسترسی به داده‌های <i>DataFrame</i>
۵۴.....	چاپ تمام مقادیر داخل دیتافریم
۵۴.....	دسترسی به داده‌های دیتافریم با استفاده از <i>iloc</i>
۵۵.....	استفاده از تکنیک <i>slice</i> یا (:)
۵۵.....	دسترسی به یک عنصر خاص دیتافریم
۵۶.....	دسترسی به داده‌های دیتافریم با استفاده از <i>loc</i>
۵۶.....	تغییر مقادیر دیتافریم
۵۷.....	حذف سطر و یا ستون از دیتافریم
۵۸.....	اعمال تغییر روی داده‌ها با استفاده از تابع <i>lambda</i>
۵۹.....	افزودن ستون به دیتافریم
۵۹.....	افزودن سطر به دیتافریم

فصل چهارم / پیش پردازش داده‌ها ۶۱

۶۲..... مدیریت داده‌های ازدست‌رفته

۶۳..... استفاده از مقدار قبلی

۶۴..... استفاده از مقدار بعدی

۶۴..... پر کردن مقادیر ازدست‌رفته بر اساس نظر یک فرد خیره

۶۵..... حذف سطر

۶۶..... حذف ستون

۶۶..... تکنیک *simpleimputer*

۶۹..... تکنیک *Group by* برای پر کردن مقادیر ازدست‌رفته

۷۰..... نرمال‌سازی داده‌ها

۷۱..... مقیاس کردن داده‌ها

۷۵..... تبدیل داده‌های *categorical* به داده‌های عددی

۷۵..... تبدیل داده‌های *categorical* بدون ترتیب به معادل عددی با استفاده از تکنیک *One-hot*

۷۸..... تبدیل داده‌های *categorical* با ترتیب به معادل عددی

۷۹..... تقسیم داده‌ها به آموزش و آزمایش

۸۰..... اعتبارسنجی متقابل

۸۱..... تکنیک‌های کاهش ابعاد

۸۱..... انتخاب ویژگی

۸۱..... روش انتخاب ویژگی *Wrapper*

۸۲..... روش انتخاب ویژگی *Filter*

۸۳..... روش انتخاب ویژگی *embedded*

۸۳..... استخراج ویژگی

۸۳..... تحلیل مؤلفه اساسی *Principal component analysis (PCA)*

فصل پنجم / رسم نمودارها و تجسم داده‌ها و آشنایی با کتابخانه‌های *SEABORN* و *MATPLOTLIB* ۸۵

۸۶..... رسم نمودارها و نمایش داده‌ها با استفاده از کتابخانه *SEABORN* و *MATPLOTLIB*

۸۶..... رسم *FIGURE* برای نمایش نمودارها

۸۷..... نمودار پراکنندگی

۸۸..... رسم *scatter plot* با *marker* های مختلف

۸۸..... رسم *scatter plot* با سایز (*s*) و رنگ‌های متفاوت (*c*)

۸۹..... رسم *scatter plot* بر اساس سایز داده‌ها

۸۹..... تعیین میزان شفافیت و رنگ لبه مارکرها

۹۰.....	تعیین بازه محور x و y
۹۱.....	تغییر برجسب‌های محور x و y
۹۲.....	قرار دادن <i>label</i> برای محور افقی، محور عمودی و عنوان برای نمودار.....
۹۴.....	رسم نمودار <i>scatter</i> با استفاده از کتابخانه <i>Seaborn</i>
۹۶.....	رسم نمودار هیستوگرام با استفاده از کتابخانه <i>Matplotlib</i>
۹۶.....	رسم نمودار هیستوگرام با استفاده از کتابخانه <i>Seaborn</i>
۹۹.....	رسم نمودار ستونی با استفاده از کتابخانه <i>Matplotlib</i>
۱۰۰.....	رسم نمودار ستونی با استفاده از کتابخانه <i>Seaborn</i>
۱۰۲.....	رسم نمودار خطی با استفاده از <i>Matplotlib</i>
۱۰۵.....	قرار دادن برجسب روی نمودار.....
۱۰۶.....	استفاده از <i>rc</i> برای تعیین مکان لیبل.....
۱۰۶.....	نمودار دایره‌ای با استفاده از <i>matplotlib</i>
۱۰۷.....	نمودار دوناتی.....
۱۰۹.....	رسم <i>subplot</i> با استفاده از <i>Matplotlib</i>
۱۱۱.....	رسم <i>subplot</i> به صورت عمودی.....
۱۱۳.....	رسم <i>subplot</i> در دو جهت افقی و عمودی.....
۱۱۴.....	رسم چند نمودار روی هم.....
۱۱۵.....	رسم نمودارهای ۳ بعدی.....
۱۱۷.....	افزودن <i>colorbar</i> به نمودار ۳ بعدی.....
۱۱۹.....	نمودار جعبه‌ای.....
۱۲۱.....	رسم نمودار جعبه‌ای با استفاده از <i>matplotlib</i>
۱۲۳.....	نمودار جعبه‌ای با استفاده از <i>seaborn</i>
۱۲۴.....	رسم نمودار جفتی.....
۱۲۵.....	ماتریس همبستگی.....

۱۲۸..... فصل ششم / یادگیری ماشین

۱۲۹.....	یادگیری ماشین.....
۱۲۹.....	یادگیری نظارت‌شده.....
۱۳۰.....	دسته‌بندی.....
۱۳۰.....	مفاهیم و معیارهای ارزیابی.....
۱۳۰.....	ماتریس درهم‌ریختگی.....
۱۳۳.....	نمودار مشخصه عملکرد.....
۱۳۴.....	بیش‌برازش.....

۱۳۴	کم برازش (<i>underfitting</i>)
۱۳۵	الگوریتم‌های دسته‌بندی
۱۳۵	الگوریتم k نزدیک‌ترین همسایه
۱۳۹	تعیین k بهینه
۱۴۰	الگوریتم ماشین بردار پشتیبان
۱۴۲	مرز تصمیم خطی
۱۴۳	مرز تصمیم غیرخطی
۱۴۴	حلقه کرنل
۱۴۴	کرنل چندجمله‌ای
۱۴۴	تابع پایه شعاعی گاوسی
۱۴۸	ماشین بردار پشتیبان با حاشیه سخت و نرم
۱۵۱	درخت تصمیم
۱۵۴	بهره اطلاعات
۱۵۴	ضریب جینی
۱۵۷	بیز
۱۵۸	بیز ساده
۱۶۲	رگرسیون
۱۶۲	رگرسیون خطی ساده
۱۶۵	رگرسیون خطی چندمتغیره
۱۶۹	رگرسیون لجستیک
۱۷۶	الگوریتم‌های بدون ناظر
۱۷۶	خوشه‌بندی
۱۷۷	روش‌های خوشه‌بندی
۱۷۷	۱- خوشه خوشه‌بندی تفکیکی
۱۸۴	۲- خوشه‌بندی سلسله‌مراتبی
۱۸۶	دندروگرام
۱۸۹	فصل هفتم / پیاده‌سازی یک مسأله دسته‌بندی به صورت مرحله به مرحله
۱۹۰	مراحل اجرا
۱۹۰	فراخوانی کتابخانه‌های مورد نیاز
۱۹۱	بارگذاری داده‌ها
۱۹۲	بررسی ارتباط بین ویژگی‌ها با استفاده از نمودار <i>pairplot</i>
۱۹۳	نمایش داده‌ها با استفاده از نمودار <i>heatmap</i>

۱۹۴ نمایش توصیف داده‌ها
۱۹۴ نمایش اطلاعات داده‌ها
۱۹۴ پیش‌پردازش داده‌ها
۱۹۵ ویژگی‌های داده‌های طبقه‌بندی‌شده را به فرمت عددی تبدیل کنید:
۱۹۵ مقیاس کردن داده‌ها
۱۹۶ ساخت مدل و استفاده از <i>Pipeline</i> و <i>GridsearchCv</i>

خط‌مشی انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب‌هایی با کیفیت عالی است که بتواند
خواسته‌های به‌روز 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 دست‌دارید تألیف "سرکار خانم نفیسه رمضان پور" است که با تلاش همکاران ما در نشر دیباگران تهران منتشر گشته و 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 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
dibagaran@mftplus.com

مقدمه مولف

سلام

این کتاب مقدمه‌ای بر مفاهیم الگوریتم‌های یادگیری ماشین می‌باشد، که نیازی به دانش قبلی در مورد یادگیری ماشین یا هوش مصنوعی¹ ندارد. در این کتاب سعی شده‌است تمامی مفاهیم به زبان ساده و به دور از محاسبات پیچیده ریاضی تحلیل شود و بیشتر بر جنبه‌های کاربردی و عملی استفاده از الگوریتم‌های یادگیری ماشین تمرکز کرده‌ایم. دوستان علاقه‌مند به حوزه آمار و ریاضیات پایه یادگیری ماشین می‌توانند به کتاب "The elements of statistical learning" نوشته Trevor Hastie مراجعه کنند. هدف این کتاب معرفی کتابخانه‌های پایتون در زمینه هوش مصنوعی می‌باشد.

کتابخانه‌های `seaborn` و `matplotlib`، `numpy`، `pandas` و `scikit-learn` در این کتاب با مثال‌های کاربردی در زمینه‌های مختلف بصری‌سازی داده‌ها، معیارهای ارزیابی الگوریتم‌ها، رسم انواع نمودارها، بررسی الگوریتم‌های یادگیری ماشین و پیش‌پردازش داده‌ها معرفی شده‌اند.

امید است این کتاب راهی باشد برای موفقیت دوستداران حوزه یادگیری ماشین و هوش مصنوعی.

نقیسه رضانی‌پور

¹ AI