



به نام خدا

یادگیری ماشین و یادگیری عمیق با زبان های پایتون و R

مؤلف:

مسعود کاویانی



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 می گیرند.

◀ عنوان کتاب: یادگیری ماشین و یادگیری عمیق با زبان های پایتون و R

◀ مولف: مسعود کاویانی

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ ویراستار: نرگس مهربد

◀ صفحه آرای: نازنین نصیری

◀ طراح جلد: داریوش فرسای

◀ نوبت چاپ: اول

◀ تاریخ نشر: ۱۴۰۱

◀ چاپ و صحافی: صدف

◀ تیراژ: ۱۰۰ جلد

◀ قیمت: ۱۸۰۰۰۰۰ ریال

◀ شابک: ۹۷۸-۶۲۲-۲۱۸-۶۳۳-۳

◀ 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-tehran.com

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

سرشناسه: کاویانی، مسعود، ۱۳۶۹- kavian,masoud
عنوان و نام پدیدآور: یادگیری ماشین و یادگیری عمیق با زبان های پایتون و R / مولف: مسعود کاویانی؛ ویراستار: نرگس مهربد.
مشخصات نشر: تهران: دیباگران تهران: ۱۴۰۱
مشخصات ظاهری: ۲۴۶ ص: مصور، جدول، نمودار
شابک: ۹۷۸-۶۲۲-۲۱۸-۶۳۳-۳
وضعیت فهرست نویسی: فیبا
موضوع: فراگیری ماشینی machine learning
موضوع: یادگیری عمیق deep learning
موضوع: داده کاوی data mining
موضوع: ساختار داده ها-الگوهای ریاضی
موضوع: data structures(computer science)-mathematical models
موضوع: پایتون(زبان برنامه نویسی کامپیوتر)
موضوع: python(computer program language)
موضوع: آ(زبان برنامه نویسی کامپیوتر)
موضوع: R(computer program language)
رده بندی کنگره: Q ۳۲۵/۵
رده بندی دیویی: ۰۰۶/۳۱
شماره کتابشناسی ملی: ۹۱۰۷۱۷۵

فهرست مطالب

۷ مقدمه ناشر

۸ مقدمه مؤلف

۹ مبانی داده کاوی و یادگیری ماشین **فصل اول**

۹ چرا داده کاوی؟ چرا یادگیری ماشین؟

۱۰ چند نمونه کاربردی از داده کاوی و یادگیری ماشین در صنعت

۱۱ تفاوت هوش مصنوعی، علم داده، داده کاوی و یادگیری ماشین

۱۲ زیرحوزه‌های داده کاوی و یادگیری ماشین

۱۳ انواع ساختارهای مسائل در داده کاوی و یادگیری ماشین

۱۴ داده‌های مستطیلی و ساختاریافته

۱۴ داده‌های تصویری

۱۵ داده‌های صوتی و سیگنال

۱۵ داده‌های متنی

۱۵ داده‌های ویدیویی

۱۵ رویکرد یادگیری نظارت شده در مقابل یادگیری غیر نظارت شده

۱۷ یادگیری دسته‌ای در مقابل یادگیری برخط

۱۸ تفاوت نمونه و متغیر

۱۹ مندولوژی و فرآیند کریسپ در داده کاوی و یادگیری ماشین

۱۹ شناخت کسب و کار

۲۰ شناخت داده‌ها

۲۱ آماده سازی داده‌ها

۲۱ مدل سازی بر روی داده‌ها

۲۱ ارزیابی مدل

۲۱ انتشار

۲۲ مقدماتی درباره‌ی جبرخطی

۲۳ ماتریس‌ها، نمونه‌ها و متغیرها

۲۵ معیارهای فاصله

۲۶ ساختار فصل‌های آینده

۲۷ این کتاب چه تفاوتی با کتاب‌های مشابه دارد؟

۲۸ پایتون و R، نصب و مقدمات **فصل دوم**

۲۸ چرا زبان پایتون، چرا زبان R؟

| | |
|---------|--|
| ۳۰..... | زبان برنامه‌نویسی پایتون |
| ۳۰..... | نصب پایتون |
| ۳۳..... | اجرای اولین برنامه در پایتون |
| ۳۹..... | آموزش مقدماتی پایتون |
| ۵۰..... | زبان برنامه‌نویسی R |
| ۵۰..... | نصب R |
| ۵۳..... | اجرای اولین برنامه در R |
| ۵۵..... | آموزش مقدماتی زبان R |
| ۶۴..... | کتابخانه‌های مورد استفاده در پایتون و R برای داده‌کاوی و یادگیری ماشین |
| ۶۴..... | لیست کتابخانه‌های پایتون |
| ۶۵..... | لیست کتابخانه‌های R |

فصل سوم شناخت و تحلیل اکتشافی داده‌ها ۶۷

| | |
|----------|--|
| ۶۷..... | مقدمه |
| ۶۸..... | بصری‌سازی داده‌ها |
| ۶۹..... | نمودار هیستوگرام |
| ۷۱..... | نمودار نقطه‌ای |
| ۷۲..... | نمودار خطی |
| ۷۳..... | نمودار میله‌ای |
| ۷۳..... | نمودار دایره‌ای |
| ۷۴..... | آمار توصیفی |
| ۷۴..... | مرکزیت داده‌ها |
| ۷۶..... | پراکندگی در داده‌ها |
| ۷۶..... | چولگی در داده‌ها |
| ۷۸..... | کشیدگی در داده‌ها |
| ۷۸..... | همبستگی در داده‌ها |
| ۸۰..... | خلاصه‌ی پنج عددی |
| ۸۱..... | مثال داده‌های فرودگاه (کد پایتون) |
| ۹۱..... | مثال داده‌های مصرف سوخت خودروها (کد R) |
| ۱۰۶..... | خلاصه فصل |

فصل چهارم پیش‌پردازش و آماده‌سازی داده‌ها ۱۰۷

| | |
|----------|---------------------------------|
| ۱۰۷..... | مقدمه |
| ۱۰۹..... | مقابله با داده‌های گم‌شده |
| ۱۱۲..... | عددی‌سازی داده‌ها |
| ۱۱۳..... | هم‌مقیاس کردن داده‌ها |

| | |
|----------|---|
| ۱۱۵..... | انتخاب ویژگی و کاهش ابعاد |
| ۱۱۷..... | تشخیص داده‌های پرت |
| ۱۱۸..... | مثال مشتریان فروشگاه اینترنتی (کد پایتون) |
| ۱۲۶..... | مثال مشتریان فروشگاه اینترنتی (کد R) |
| ۱۲۹..... | خلاصه فصل |

فصل پنجم یادگیری نظارت‌شده: طبقه‌بندی و رگرسیون ۱۳۱

| | |
|----------|--|
| ۱۳۱..... | مقدمه |
| ۱۳۵..... | الگوریتم‌های طبقه‌بندی و رگرسیون |
| ۱۳۵..... | الگوریتم نزدیک‌ترین همسایه‌ها |
| ۱۳۷..... | الگوریتم ماشین بردار پشتیبان |
| ۱۴۰..... | الگوریتم درخت تصمیم |
| ۱۴۲..... | الگوریتم جنگل تصادفی |
| ۱۴۳..... | الگوریتم آداپوست |
| ۱۴۴..... | الگوریتم ایکس‌جی‌بوست و کت‌بوست |
| ۱۴۴..... | ارزیابی الگوریتم‌های طبقه‌بندی و رگرسیون |
| ۱۴۶..... | روش‌های ارزیابی الگوریتم‌های طبقه‌بندی |
| ۱۴۸..... | روش‌های ارزیابی الگوریتم‌های رگرسیون |
| ۱۵۰..... | بیش‌برازش و کم‌برازش |
| ۱۵۱..... | روش‌های افزایش دقت در الگوریتم‌های طبقه‌بندی و رگرسیون |
| ۱۵۲..... | مثال مجموعه‌ی داده‌ی طبقه‌بندی گل‌های زنبق (کد پایتون) |
| ۱۵۵..... | مثال پردازش تصویر تشخیص اعداد دست‌نویس (کد پایتون) |
| ۱۶۷..... | مثال تشخیص بیماری قلبی (کد پایتون) |
| ۱۷۵..... | مثال پیش‌بینی قیمت مسکن (کد پایتون) |
| ۱۸۰..... | مثال تشخیص سرطان (کد R) |
| ۱۸۵..... | مثال تخمین خسارت حادثه شرکت بیمه (کد R) |
| ۱۸۹..... | خلاصه‌ی فصل |

فصل ششم یادگیری غیرنظارت‌شده: خوشه‌بندی ۱۹۱

| | |
|----------|---|
| ۱۹۱..... | مقدمه |
| ۱۹۳..... | الگوریتم‌های خوشه‌بندی |
| ۱۹۳..... | الگوریتم K میانگین |
| ۱۹۶..... | الگوریتم خوشه‌بندی مبتنی بر غلظت DBSCAN |
| ۱۹۸..... | الگوریتم تغییر میانگین |
| ۲۰۰..... | ارزیابی الگوریتم‌های خوشه‌بندی |
| ۲۰۰..... | روش جمع مربعات درون خوشه‌ای |

| | |
|----------|--|
| ۲۰۰..... | روش نیمرخ..... |
| ۲۰۱..... | مثال خوشه‌بندی مشتریان فروشگاه اینترنتی (کد پایتون)..... |
| ۲۰۸..... | خوشه‌بندی کاربران شبکه‌های اجتماعی (کد R)..... |
| ۲۱۲..... | خلاصه‌ی فصل..... |

فصل هفتم شبکه‌های عصبی مصنوعی و یادگیری عمیق..... ۲۱۳

| | |
|----------|---|
| ۲۱۳..... | مقدمه..... |
| ۲۱۷..... | الگوریتم‌های شبکه‌های عصبی و یادگیری عمیق..... |
| ۲۱۸..... | پرسپترون چندلایه و شبکه‌های پیش‌خور..... |
| ۲۱۸..... | شبکه‌های باور عمیق و ماشین بولتزن محدودشده..... |
| ۲۱۹..... | شبکه‌های عصبی پیچشی..... |
| ۲۲۰..... | شبکه‌های عصبی بازگشتی..... |
| ۲۲۲..... | توابع فعال‌سازی..... |
| ۲۲۳..... | لایه‌ی حذف تصادفی..... |
| ۲۲۴..... | مثال پیش‌بینی بیماری دیابت (کد پایتون)..... |
| ۲۲۷..... | مثال پیش‌بینی بیماری ریه از تصاویر اشعه ایکس (کد پایتون)..... |
| ۲۳۴..... | مثال پیش‌بینی قیمت سهام (کد پایتون)..... |
| ۲۳۹..... | مثال تشخیص کیفیت قطعات تولیدی (کد R)..... |
| ۲۴۳..... | مثال تشخیص سگ در یک تصویر (کد R)..... |
| ۲۴۶..... | خلاصه فصل..... |

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 دست دارید تألیف "جناب آقای مسعود کاویانی" 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 مؤلف

کتاب‌های زیادی در حوزه‌ی هوش مصنوعی، داده‌کاوی، یادگیری ماشین و یادگیری عمیق وجود دارند که هر کدام سعی کرده‌اند از زاویه‌ی دید مؤلف اثر، مفاهیم موجود در این حوزه را به خواننده منتقل کنند. در میان این کتاب‌ها، کتابی که بتواند به صورت ساده و تا حدودی مختصر و مبتنی بر مثال‌های کاربردی، مبانی و مفاهیم این حوزه را با حداقل پیش‌نیاز ارائه کند یافت نکردم. از این رو بر آن شدم تا با جمع‌آوری و پیاده‌سازی مثال‌های کاربردی و متفاوت در یک ساختار مفهومی، مبانی و بخش‌های اصلی در یادگیری ماشین و یادگیری عمیق را به زبان پارسی در یک کتاب ارائه دهم.

این کتاب برای دانشجویان و خوانندگانی مناسب است که می‌خواهند علاوه بر مبانی و مفاهیم، بتوانند با استفاده از یک زبان برنامه‌نویسی مانند پایتون یا R، به حل مسائل دنیای واقعی بپردازند. از همین جهت شرح و بسط الگوریتم‌ها و روش‌ها را تا حد امکان خلاصه کرده و بیشتر بر جنبه‌ی درک شهودی در بیان الگوریتم‌ها کوشش نموده‌ام.

این کتاب را به همسرم تقدیم می‌کنم. همسری که در تمامی مراحل زندگی پشتیبان و همراه من بوده است. همچنین تقدیم می‌کنم به پدر و مادرم که در سختی‌ها و مشکلات زندگی یاور و پشتیبان من بوده‌اند.