

دیباگران تهران

مؤسسه فرهنگی هنری

دیباگران تهران

مؤسسه فرهنگی هنری

به نام خدا

یادگیری ماشین

با زبان برنامه نویسی

پایتون

مؤلفان:

مهندس طاهره اسماعیلی ابهریان

مهندس محسن علی مرادی



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

عنوان کتاب: **یادگیری ماشین با زبان برنامه نویسی پایتون**

مولفان: مهندس طاهره اسمعیلی ابهریان

مهندس محسن علی مرادی

ناشر: موسسه فرهنگی هنری دیباگران تهران

ویراستار: زهرا خانیانی

صفحه آرایی: نازنین نصیری

طراح جلد: داریوش فرسایی

نوبت چاپ: اول

تاریخ نشر: ۱۳۹۸

چاپ و صحافی: صدف

تیراژ: ۱۰۰۰ جلد

قیمت: ۵۸۰۰۰ ریال

شابک: ۹۷۸-۶۲۲-۲۱۸-۲۱۵-۱

نشانی واحد فروش: تهران، میدان انقلاب،
خ کارگر جنوبی، روبروی پاساز مهستان،
پلاک ۱۲۵۱

تلفن: ۰۲۰-۸۵۱۱۱-۶۶۴۱۰۰۴۶

فروشگاههای اینترنتی دیباگران تهران:

WWW.MFTBOOK.IR

www.dibbook.ir

www.dibagaran Tehran.com

نشانی تلگرام: [@mftbook](https://t.me/mftbook)

هر کتاب دیباگران، یک فرصت جدید شغلی.

هرگوشی همراه، یک فروشگاه کتاب دیباگران تهران.

از طریق سایتها و اپ دیباگران، در هر جای ایران به کتابهای ما دسترسی دارید.

سرشناسه: اسمعیلی ابهریان، طاهره، ۱۳۶۵-

عنوان و نام پدیدآور: یادگیری ماشین با زبان برنامه نویسی پایتون / مولفان: طاهره اسمعیلی ابهریان، محسن علی مرادی.

مشخصات نشر: تهران: دیباگران تهران: ۱۳۹۸

مشخصات ظاهری: ۱۸۰ ص: مصور،
شابک: ۹۷۸-۶۲۲-۲۱۸-۲۱۵-۱

وضعیت فهرست نویسی: فیبا

موضوع: پایتون (زبان برنامه نویسی کامپیوتر)
Python (computer program language)

شناسه افزوده: علی مرادی، محسن، ۱۳۶۵-

رده بندی کنگره: QA ۷۶/۷۳

رده بندی دیوبی: ۰۵/۱۳۲

شماره کتابشناسی ملی: ۵۹۴۱۰۵۶

فهرست مطالب

۶	مقدمه ناشر
۷	مقدمه مؤلفان
۸	پیشگفتار

فصل اول

۹	مفاهیم یادگیری ماشین
۱۰	۱- یادگیری ماشین چیست؟
۱۱	۲- مدل یادگیری ماشین
۱۱	۳- چهار رویکرد اصلی یادگیری ماشین
۱۲	۱-۳- یادگیری با ناظر
۱۴	۱-۲-۳- یادگیری بدون ناظر
۱۶	۱-۳-۳- یادگیری نیمه نظارتی
۱۶	۱-۴-۳- یادگیری تقویتی
۱۷	۱-۴- معرفی مجموعه داده
۱۸	۱-۵- معرفی روند کلی یک سیستم یادگیری ماشین
۱۸	۱-۵- ۱- پیش‌پردازش داده
۲۱	۱-۵- ۲- تقسیم مجموعه داده
۲۲	۱-۵- ۳- انتخاب و آموزش یک مدل یادگیری
۲۳	۱-۵- ۴- ارزیابی مدل و پیش‌بینی داده‌هایی که تاکنون با مدل دیده نشده
۲۳	۱-۶- شروع استفاده از پایتون در یادگیری ماشین
۲۳	۱-۶- ۱- نصب پایتون

فصل دوم

۲۶.....	دسته‌بندی داده با پایتون.....
۲۷.....	۱-۱-۲- تا نزدیک‌ترین همسایه (KNN).....
۳۵.....	۱-۱-۲- استفاده از ماثول knn برای دسته‌بند
۳۸.....	۲-۲- شبکه‌های عصبی مصنوعی
۳۸.....	۱-۲-۲- شبکه عصبی پرسپترون تک لایه
۴۶.....	۲-۲-۲- شبکه عصبی پرسپترون چندلایه (MLP)
۵۳.....	۳-۲-۲- شبکه عصبی پس انتشار (BP)
۷۲.....	۳-۳-۲- ماشین بردار پشتیبان (SVM)
۷۹.....	۱-۳-۲- فراتر از مرزهای خطی: هسته SVM
۸۴.....	۲-۳-۲- تنظیم SVM: نرم کردن حاشیه‌ها
۸۶.....	۴-۲- قضیه بیز
۸۷.....	۱-۴-۲- طبقه‌بند بیز
۱۰۱.....	۵-۲- درخت تصمیم
۱۰۳.....	۱-۵-۲- محاسبه بیشترین کسب اطلاعات (IG)

فصل سوم

۱۰۸.....	خوش‌بندی داده با پایتون.....
۱۰۹.....	۱-۳- الگوریتم K-MEANS
۱۱۵.....	۲-۳- الگوریتم FUZZY C-MEANS
۱۱۸.....	۳-۳- الگوریتم سلسه‌مراتبی

فصل چهارم

۱۲۴.....	پیش‌پردازش داده
۱۲۵.....	۱-۴- مدیریت «اطلاعات گم شده» در مجموعه داده
۱۲۸.....	۲-۴- مدیریت داده‌های دسته‌ای
۱۳۲.....	۳-۴- تقسیم یک مجموعه داده به بخش‌های آموزش و آزمون
۱۳۴.....	۴-۴- یکسان کردن مقیاس ویژگی‌ها

فصل پنجم

۱۳۷..... کاهش ابعاد داده
۱۳۸..... ۱-۵- مشکل بیش برآش
۱۳۸..... ۲-۵- روش‌های انتخاب ویژگی
۱۴۰..... ۱-۲-۵- انتخاب ویژگی با جنگل‌های تصادفی
۱۴۳..... ۲-۲-۵- انتخاب ویژگی با روش انتخاب پشت سرهم (SBS)
۱۴۸..... ۳-۵- استخراج ویژگی
۱۴۸..... ۱-۳-۵- تحلیل مؤلفه اصلی (PCA)
۱۵۸..... ۲-۳-۵- تحلیل تفکیک‌کننده خطی (LDA)

فصل ششم

۱۶۹..... ارزیابی نهایی مدل
۱۷۱..... ۱-۶- روش‌های اعتبارسنجی
۱۷۱..... ۱-۶-۱- روش K-Fold
۱۷۱..... ۲-۱-۶- روش Leave-One-Out
۱۷۱..... ۳-۱-۶- روش Hold-out
۱۷۲..... ۴-۱-۶- روش جایگزینی مجدد
۱۷۲..... ۵-۱-۶- روش جایگشت تصادفی
۱۷۳..... ۶-۲- استفاده از PIPELINE در پایتون
۱۷۶..... منابع

مقدمه ناشر

خط میشی کیفیت انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب های است که بتواند خواسته های بر روز جامعه فرهنگی و علمی کشور را تا حد امکان پوشش دهد. هر کتاب دیباگران تهران، یک فرصت جدید شغلی

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 دانشگاهی، علوم پایه و به ویژه علوم کامپیوتر و انفورماتیک گامهایی هرچند کوچک برداشته و در انجام رسالتی که بر عهده داریم، مؤثر واقع شویم.

گسترده‌گی علوم و 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 روشن می‌نماید. در این راستا، واحد انتشارات مؤسسه فرهنگی هنری دیباگران تهران با همکاری جمعی از اساتید، مؤلفان، مترجمان، متخصصان، پژوهشگران، محققان و نیز پرستل ورزیده و ماهر در زمینه امور نشر درصد هستند تا با تلاش‌های مستمر خود برای رفع کمبودها و نیازهای موجود، منابعی پُربار، معترض و با کیفیت مناسب در اختیار علاقمندان قرار دهند.

کتابی که در دست دارید با همت "مهندسان طاهره اسماعیلی ابهریان - محسن علی مرادی " و تلاش جمعی از همکاران انتشارات میسر گشته که شایسته است از یکایک این گرامیان تشکر و قدردانی کنیم.

کارشناسی و نظارت بر محتوا: زهرو قزلباش

در خاتمه ضمن سپاسگزاری از شما دانش‌پژوه گرامی درخواست می‌نماید با مراجعه به آدرس dibagaran.mft.info (ارتباط با مشتری) فرم نظرسنجی را برای کتابی که در دست دارید تکمیل و ارسال نموده، انتشارات دیباگران تهران را که جلب رضایت و وفاداری مشتریان را هدف خود می‌داند، یاری فرمایید.

امیدواریم همواره بهتر از گذشته خدمات و محصولات خود را تقدیم حضورتان نماییم.

مدیر انتشارات

مؤسسه فرهنگی هنری دیباگران تهران
bookmarket@mft.info

به نام خداوند جان و خرد

کریم بر قرآن دیده بر نگذرد

مقدمه مؤلفان

خداوند را بسیار شاکریم که برای ما فرصت تهیء کتاب یادگیری ماشین با پایتون را به زبان فارسی فراهم کرد. تمام تلاش خود را به کار بستیم تا مباحث اصلی یادگیری ماشین را به همراه پیاده سازی آنها با زبان پایتون در این کتاب به گونه ای جمع آوری کنیم که علاوه بر مکفی بودن توضیحات الگوریتم ها، برای خوانندگان محترم خسته کننده نباشد. از آنجایی که سیستم های یادگیری ماشین به صورت گسترده ای در تمام رشته ها در حال استفاده است، سعی ما بر این شد تا چندان وارد مباحث پیچیده محاسبات و اثبات ریاضی الگوریتم ها نشویم تا این کتاب برای طیف گسترده ای از علاقه مندان به یادگیری ماشین قابل استفاده باشد.

چنانچه خوانندگان عزیز پیشنهادات یا انتقاداتی بر این نوشتار داشته باشند، سپاسگزار خواهیم بود از طریق آدرس الکترونیکی زیر، برای پر کردن آثار بعدی، ما را مطلع کنند.

طاهره اسماعیلی ابهریان، محسن علی مرادی

Tahereh.abharian@gmail.com

msn.alimoradi@gmail.com

پیشگفتار

درصورتی که در هر حوزه‌ای از فناوری اطلاعات و نرم‌افزار مشغول به فعالیت هستید، یا در حوزه مهندسی برق، مکانیک، مهندسی پزشکی، شیمی، بیوشیمی، بیوفیزیک، بیوانفورماتیک و ژنتیک فعالیت می‌کنید، یا حتی اگر یک داروساز یا پزشک هستید، این کتاب قطعاً به شما کمک خواهد کرد تا درک بهتری از یادگیری یک سیستم هوشمند داشته باشید.

پردازش تصاویر دیجیتال از طریق الگوریتم‌های یادگیری ماشین و تحلیل‌های آماری، کاربرد گسترده‌ای در حوزه‌های مختلفی مانند تشخیص توده‌های سرطانی، تشخیص هویت افراد با استفاده از تصاویر اثرازنگشت و عنیبه، تشخیص پلاک خودرو، مسیریابی ربات‌ها در صنایع فضایی و دفاعی و طراحی ماشین‌های بدون سرنشین و... دارد.

پیش‌بینی بیماری‌ها از طریق پردازش فاکتورهای خونی و ژنتیکی، کمک شایانی به سلامت جامعه جهانی خواهد کرد. پژوهشگران حوزه بیوانفورماتیک در تلاش‌اند تا با کمک داروسازان و پزشکان، با استفاده از داده‌هایی که هر روز به سرعت تولیدشان افروده می‌شود، درمان بسیاری از بیماری‌ها را بیابند.

هدف این کتاب این است که یادگیری ماشین را برای دانش‌پژوهان به‌گونه‌ای کاربردی معرفی کند تا خوانندگان عزیز با داشتن اطلاعات پایه‌ای از برنامه‌نویسی و زبان پایتون بتوانند در حوزه فعالیت خود از این الگوریتم‌ها استفاده کنند.

اکنون فرصت را غنیمت می‌شماریم تا زحمات بی‌دریغ پدران و مادران عزیzman را با جملاتی هرچند کوتاه ارج نهیم. پدر و مادر عزیzman، شما قدم‌به‌قدم چگونه زیستن را در مکتب عشق به ما آموختید. در مقابل عظمت و شکوه شما ما را نه توان سپاس است و نه کلام وصف. به پاس محبت‌های بی‌دریغتان، که هرگز فروکش نمی‌کند، این کتاب را به شما پدر و مادر عزیzman تقدیم می‌کنیم.

تقدیم به پدر و مادر عزیzman و فرزند دلبندمان

طاهره اسماعیلی ابهریان، محسن علی مرادی