



به نام خدا



ارائه راهکار برای تمام مسائل یادگیری ماشین

مؤلفان:

دکتر محسن یزدی نژاد (دکترای هوش مصنوعی)

دکتر اسماعیل زاهدی (دکترای هوش مصنوعی)

دکتر محمد رضایپور (عضو هیئت علمی دانشگاه و پژوهشگر پسا دکترا)



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: ارائه راهکار برای تمام مسائل یادگیری ماشین

◀ مولفان: دکتر محسن یزدی نژاد - دکتر اسماعیل زاهدی - دکتر محمد رضا پور

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ ویراستار: نرگس مهربد

◀ صفحه آرای: شبنم هاشم زاده

◀ طراح جلد: داریوش فرسای

◀ نوبت چاپ: اول

◀ تاریخ نشر: ۱۴۰۰

◀ چاپ و صحافی: درج عقیق

◀ تیراژ: ۱۰۰ جلد

◀ قیمت: ۱۰۹۰۰۰۰ ریال

◀ شابک: ۹۷۸-۶۲۲-۲۱۸-۴۵۹-۹

◀ 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

سرشناسه: یزدی نژاد، محسن، ۱۳۶۲-

عنوان و نام پدیدآور: ارائه راهکار برای تمام مسائل یادگیری ماشین/مولفان: محسن یزدی نژاد، اسماعیل زاهدی، محمد رضا پور.

مشخصات نشر: تهران: دیباگران تهران: ۱۴۰۰

مشخصات ظاهری: ۲۷۸ ص: مصور،

شابک: ۹-۴۵۹-۲۱۸-۶۲۲-۹۷۸

وضعیت فهرست نویسی: فیبا یادداشت: کتابنامه: ص ۲۷۷-۲۷۸

موضوع: فراگیری ماشینی Machine Learning

موضوع: پایتون (زبان برنامه نویسی کامپیوتر)

موضوع: Python (computer program language)

شناسه افزوده: زاهدی، اسماعیل، ۱۳۶۸-

شناسه افزوده: رضا پور، محمد، ۱۳۵۷-

رده بندی کنگره: ۳۲۵/۵ QC

رده بندی دیویی: ۰۰۶/۳۱

شماره کتابشناسی ملی: ۸۵۴۰۸۲۳

نشانی اینستاگرام دیبا dibagaran_publishing نشانی تلگرام: @mftbook

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

۸	مقدمه نصب و تنظیم پیش‌نیازهای کدنویسی
۱۰	فصل اول یادگیری با نظارت در مقایسه با یادگیری بدون نظارت
۱۸	فصل دوم اعتبارسنجی متقابل
۳۳	فصل سوم معیارهای ارزیابی
۶۹	فصل چهارم سامان‌دهی پروژه‌های یادگیری ماشین
۸۰	فصل پنجم مروری بر متغیرهای گسسته
۱۲۷	فصل ششم مهندسی ویژگی
۱۴۰	فصل هفتم انتخاب ویژگی

فصل هشتم

۱۵۴ بهینه‌سازی هایپر پارامترها

فصل نهم

۱۶۹ مروری بر دسته‌بندی و جداسازی تصاویر دیجیتال

فصل دهم

۲۱۷ مروری بر رگرسیون یا دسته‌بندی متون

فصل یازدهم

۲۵۵ مروری بر روش‌های تجمیع و stacking

فصل دوازدهم

۲۶۴ مروری بر استقرار و به کارگیری مدل به صورت عملی

۲۷۷ منابع و مآخذ

خط‌مشی انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب‌هایی با کیفیت عالی است که بتواند
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 "آقایان دکتر محسن یزدی نژاد- دکتر اسماعیل زاهدی- دکتر محمد رضاپور" و همکاری سرکار خانم‌ها ساره هرمان-مهسا محسن بیگی 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
dibagaran@mftplus.com

مقدمه مولف

وضعیت حاضر که در سال ۲۰۲۱ میلادی یا ۱۴۰۰ خودمان به سر می‌بریم، یک وضعیت بحرانی از لحاظ مقدار و تنوع داده‌هایی است که در هر لحظه تولید شده و متأسفانه بخش قابل توجهی از آنها نیز پردازش نمی‌شوند! این نگرانی از اینروستکه با تحلیل صحیح داده‌ها می‌توان از روندها و آنگوهای موجود در آنها مطلع شد، به امر حیاتی «پیش‌بینی» تا حدّ زیادی دست یافت و از بازتکرار مسیرهای پرهزینه یا منجر به شکست پیشگیری نمود.

علت «عدم تحلیل کافی داده‌ها» در هر موقعیتی مختص به همان جاست و نمی‌توان با یک پاراگراف به آن پاسخ داد، اما آنچه می‌توان در زمره‌ی وجه اشتراک این وضعیت نام‌برد هم به‌لحاظ شکلی و ساختاری است و هم به‌دلیل محتوایی و دانشی؛ از لحاظ شکلی، عدم تجمیع داده‌ها و سامانه‌ها، موجب شده تا بخش زیادی از دانسته‌های بالقوه‌ی مجریان، بالفعل نشوند و برخورد جزیره‌ای در معماری پیاده‌سازی داده‌ها – حداقل توسط سازمان‌های مختلف – موجب شده تا بخش‌های مختلف از پدیده‌ی «یکپارچه‌سازی داده‌ها» محروم باشند. به لحاظ محتوایی نیز کم نیستند محیط‌هایی که با افراد متبحری مواجه هستند که عمده‌ی توانمندی فنی ایشان به مدیریت داده‌های ساختاریافته^۱ محدود بوده و در زمینه‌ی «تحلیل داده‌هایی مانند متن و تصویر» حتی از آشنایی سطح متوسط هم برخوردار نیستند؛ بگذریم که چقدر سرفصل‌های اصلی رشته‌های مهندسی در دانشگاه‌ها، در این زمینه قصور داشته‌اند و از توجه عملی به نیازمندی فضای کسب و کار عقب مانده‌اند.

شایان گفتن اینکه تأکید می‌کنیم مخاطب این کتاب، افرادی هستند که می‌خواهند سطح «حرفه‌ای» یادگیری ماشین را طی کنند و آشنایی مقدماتی با الگوریتم‌های یادگیری ماشین و همچنین زبان برنامه‌نویسی پایتون را دارند؛ عزیزی که در ابتدای این مسیر قرار دارند، می‌توانند از سه کتاب دیگرمان با عناوین «هم‌اکنون، داده‌کاوی»، «علم داده در پایتون – مفاهیم و کاربردها» و «آمار کاربردی در علم داده» بهره‌مند شده و سپس نکات کتاب حاضر را فرا بگیرند. از منظر زبان برنامه‌نویسی، تمام قسمت‌های این کتاب بر اساس کد پایتون بوده، لذا این کتاب کاملاً جنبه عملی دارد و می‌توان از آنها در پروژه‌های عملی استفاده کرد.

شما با داشتن آشنایی مقدماتی با مواردی که گفتیم و با مطالعه مطالب این کتاب می‌توانید توانمندی‌های خود را در این زمینه ارتقا دهید و برای شرکت در چالش‌های کگل و همچنین محیط کار آماده شوید. علاوه بر این، بتوانید

^۱ Structured

با چنین دغدغه‌هایی شروع به نوشتن کتاب حاضر کردیم و با توجه به سطح پیشرفته‌ی کتاب، برای آنکه وقت خواننده‌ی حرفه‌ای را رعایت کرده باشیم هم به تجاربی که از سالها تدریس دوره‌های مربوطه به‌دست آمده بود توجه کردیم و هم تلاش کردیم از متون مطرح در سطح دنیا - نظیر نوشته‌های¹ Abhishek Thakur در کتاب APPROACHING (ALMOST) ANY MACHINE LEARNING PROBLEM - استفاده نماییم؛ همان‌طور که از اسم کتاب پیداست، آخرین مسائل یادگیری ماشین در آن بررسی شده است.

بدین منظور از ارزیابی‌های محیط کگل - که زیرمجموعه‌ی ای از گوگل و یک انجمن برخط (آنلاین) از دانشمندان علوم داده‌ها و نیز تمرین‌کنندگان یادگیری ماشین است - استفاده نمودیم؛ با توجه به نزدیک بودن ارزیابی چالش - های کگل به ارزیابی مسائل در دنیای واقعی، افراد موفق در این انجمن جزء بزرگترین و صاحب سبک‌ترین دانشمندان علم داده در جهان محسوب می‌شوند و آقای آبیشک تاکور، یکی از شناخته شده‌ترین‌ها در این حوزه می‌باشد. البته مواردی را که ضروری بودند ولی در نوشته‌های آبیشک به آنها پرداخته نشده بود در این متن افزوده‌ایم و در نسخه‌های بعدی هم موارد تکمیلی بیشتری - حسب نیازهای زمان چاپ آن - اضافه خواهند شد. برای آنکه عنوان کتاب (یادگیری تمام مسائل) در عمق دانش خوانندگان محقق شود، شما را به خواندن تمام کتاب پیشنهاد می‌کنیم، هرچند مطالب طوری تنظیم شده که اگر به یکباره وقت نکنید تمام کتاب را بخوانید باز هم همان فصل برایتان مفید باشد.

در پایان ضمن تشکر از ناشر محترم، از تمامی خوانندگان گرامی - اعم از اساتید، دانشجویان و مهارت‌ورزان عزیز - درخواست دارد نکات سازنده یا هر سؤالی که درباره کتاب دارند را به نویسندگان منعکس نمایند. ایام انتشار کتاب، مصادف است با ارتحال دانشمند ذوفنون، علامه حسن‌زاده‌ی آملی^(ره) که ضمن اهداء کتاب به روح مقدس ایشان، حسن ختام این دیباچه از اشعار منتسب به آن بزرگوار تقدیم می‌نمائیم:

اضطرب يضطرب اضطراب
التهب يلهب التهبا
انشعب ينشعب انشعاب
انسحب ينسحب انسحاب
انحلب ينحلب انحلاب
اقترب يقترب اقتراب

همچو گیاه لب آب روان
آتش عشق است که در اصل و فرع
نور خدایست که در شرق و غرب
آب حیاتست که در جزء و کل
از سر شوق است که اشک بصر
آملیا! موعد پیک اجل

¹ آبیشک تاکور اولین دانشمند علم داده در جهان است که در سایت کگل در هر چهار زمینه (datasets, competitions, discussions و notebooks) به عنوان استاد بزرگی دست یافته است.