

به نام خدا

---

# مهندسی یادگیری ماشین

---

نویسنده: اندری بورکوف

www.ketab.ir

مترجمان:

◀ دکتر مهدی اسماعیلی

◀ دکتر سیدمهدی وحیدی پور

# مهندسی یادگیری ماشین

مترجمان: دکتر مهدی اسماعیلی، دکتر سیدمهدی وحیدی پور

طراحی جلد و صفحه‌آرایی: همتا بیداریان

ناشر: انتشارات آتی‌نگر

ناشر همکار: انتشارات وینا

چاپ اول، ۱۴۰۰

شمارگان: ۲۰۰ نسخه

قیمت: ۹۸۰,۰۰۰ ریال

شابک: ۹۷۸-۶۲۲-۷۵۷۱-۲۶-۴

ISBN: 978-622-7571-26-4

حق چاپ برای انتشارات آتی‌نگر محفوظ است.

نشانی دفتر فروش: خیابان خمارآده جنوبی، روبه‌روی کوچه رشتچی، پلاک ۱۴۴، واحد ۱

تلفن: ۶۶۵۶۵۳۳۶-۸

www.ati-negar.com \* info@ati-negar.com



سرشناسه: بورکوف، اندری، Burkov, Andriy

مهندسی یادگیری ماشین / نویسنده: اندری بورکوف؛ مترجمان: مهدی اسماعیلی، سیدمهدی وحیدی پور

تهران: آتی‌نگر، وینا ۱۴۰۰

۳۶۰ ص.: مصور، جدول، نمودار.

ISBN: 978-622-7571-26-4

فیبا.

یادداشت: عنوان اصلی کتاب: Machine Learning Engineering

موضوع: فراگیری ماشینی -- مهندسی -- ساختار داده‌ها -- الگوهای ریاضی

موضوع: Data structures (computer science) -- Mathematical models -- Engineering -- Machine learning

شناسه‌افزوده: اسماعیلی، مهدی، ۱۳۵۰- مترجم

شناسه‌افزوده: وحیدی پور، سیدمهدی، ۱۳۵۶- مترجم

شناسه‌افزوده: بیداریان، همتا، ۱۳۶۱- گرافیکست

رده‌بندی کنگره

رده‌بندی دیویی

شماره کتابشناسی ملی

QC۳۲۵/۵

۰۰۶/۳۱

۸۴۷۶۱۷۳

# فهرست مطالب

پیشگفتار

۷

## فصل اول: مقدمه

۹

- ۱-۱ نشانه گذاری و تعریف ها ..... ۹
- ۲-۱ یادگیری ماشین چیست؟ ..... ۱۲
- ۳-۱ داده ها و یادگیری ماشین ..... ۱۶
- ۴-۱ چه موقع از یادگیری ماشین استفاده می شود ..... ۲۳
- ۵-۱ چه موقع از یادگیری ماشین استفاده نمی شود ..... ۲۷
- ۶-۱ مهندسی یادگیری ماشین چیست ..... ۲۸
- ۷-۱ چرخه حیات پروژه های یادگیری ماشین ..... ۲۹
- ۸-۱ خلاصه ..... ۳۰

## فصل دوم: قبل از این که پروژه شروع شود

۳۳

- ۱-۲ اولویت بندی پروژه های یادگیری ماشین ..... ۳۳
- ۲-۲ تخمین پیچیدگی یک پروژه یادگیری ماشین ..... ۳۶
- ۳-۲ تعریف هدف یک پروژه یادگیری ماشین ..... ۳۸
- ۴-۲ ساختار یک تیم یادگیری ماشین ..... ۴۰
- ۵-۲ چرا پروژه های یادگیری ماشین شکست می خورند ..... ۴۳
- ۶-۲ خلاصه ..... ۴۷

## فصل سوم: جمع آوری و آماده سازی داده ها

۴۹

- ۱-۳ پرسش هایی درباره داده ها ..... ۴۹
- ۲-۳ مشکلات رایج با داده ها ..... ۵۷
- ۳-۳ داده های خوب چیست ..... ۷۳
- ۴-۳ کار با داده های تعاملی ..... ۷۷
- ۵-۳ دلایل نشت داده ها ..... ۷۸
- ۶-۳ افزایش داده ها ..... ۸۱
- ۷-۳ کار با صفات خاصه ناموجود ..... ۸۴

|      |                                      |     |
|------|--------------------------------------|-----|
| ۸-۳  | داده‌افزایی.....                     | ۸۶  |
| ۹-۳  | داده‌های نامتعادل.....               | ۹۰  |
| ۱۰-۳ | استراتژی‌های نمونه‌گیری داده‌ها..... | ۹۴  |
| ۱۱-۳ | ذخیره داده‌ها.....                   | ۹۷  |
| ۱۲-۳ | تجربه کار با داده‌ها.....            | ۱۰۵ |
| ۱۳-۳ | خلاصه.....                           | ۱۰۶ |

## فصل چهارم: مهندسی ویژگی

|      |                                    |     |
|------|------------------------------------|-----|
| ۱-۴  | چرا ویژگی‌ها مهندسی می‌شوند.....   | ۱۰۹ |
| ۲-۴  | چگونه ویژگی‌ها مهندسی می‌شوند..... | ۱۱۱ |
| ۳-۴  | ویژگی‌های پشته‌ای.....             | ۱۲۸ |
| ۴-۴  | خصوصیات ویژگی‌های خوب.....         | ۱۳۰ |
| ۵-۴  | انتخاب ویژگی.....                  | ۱۳۳ |
| ۶-۴  | ویژگی‌های ساختگی.....              | ۱۳۹ |
| ۷-۴  | یادگیری ویژگی‌ها از داده‌ها.....   | ۱۴۵ |
| ۸-۴  | کاهش ابعاد.....                    | ۱۵۱ |
| ۹-۴  | مقیاس‌گذاری ویژگی‌ها.....          | ۱۵۲ |
| ۱۰-۴ | نشت داده‌ها در مهندسی ویژگی.....   | ۱۵۵ |
| ۱۱-۴ | ذخیره و مستندسازی ویژگی‌ها.....    | ۱۵۶ |
| ۱۲-۴ | پهروش‌های مهندسی ویژگی.....        | ۱۶۰ |
| ۱۳-۴ | خلاصه.....                         | ۱۶۵ |

## فصل پنجم: آموزش مدل با ناظر

|      |                                        |     |
|------|----------------------------------------|-----|
| ۱-۵  | قبل از شروع کار روی مدل.....           | ۱۷۰ |
| ۲-۵  | نمایش برجسب‌ها برای یادگیری ماشین..... | ۱۷۶ |
| ۳-۵  | انتخاب الگوریتم یادگیری.....           | ۱۷۸ |
| ۴-۵  | ساخت یک خط لوله.....                   | ۱۸۳ |
| ۵-۵  | ارزیابی عملکرد مدل.....                | ۱۸۴ |
| ۶-۵  | تنظیم اُپرِپارامترها.....              | ۱۹۷ |
| ۷-۵  | آموزش مدل کم‌عمق.....                  | ۲۰۲ |
| ۸-۵  | موازنه بایاس-واریانس.....              | ۲۰۵ |
| ۹-۵  | تنظیم.....                             | ۲۰۹ |
| ۱۰-۵ | خلاصه.....                             | ۲۱۱ |

## فصل ششم: آموزش مدل با ناظر

۲۱۵

|     |                                     |
|-----|-------------------------------------|
| ۲۱۵ | ۱-۶ استراتژی آموزش مدل عمیق         |
| ۲۳۷ | ۲-۶ مدل های پشته ای                 |
| ۲۴۰ | ۳-۶ مواجهه با تغییر توزیع           |
| ۲۴۲ | ۴-۶ مدیریت مجموعه داده های نامتعادل |
| ۲۴۵ | ۵-۶ کالیبره کردن مدل                |
| ۲۴۷ | ۶-۶ غیب یابی و تحلیل خطا            |
| ۲۵۷ | ۷-۶ به روش ها                       |
| ۲۶۵ | ۸-۶ خلاصه                           |

## فصل هفتم: ارزیابی مدل

۲۶۹

|     |                                         |
|-----|-----------------------------------------|
| ۲۷۱ | ۱-۷ ارزیابی آفلاین و آنلاین             |
| ۲۷۳ | ۲-۷ آزمون A/B                           |
| ۲۸۰ | ۳-۷ راهزن چند دست                       |
| ۲۸۴ | ۴-۷ کرانه های آماری مربوط به عملکرد مدل |
| ۲۸۸ | ۵-۷ ارزیابی کفایت مجموعه آزمایشی        |
| ۲۹۰ | ۶-۷ ارزیابی خصوصیات مدل                 |
| ۲۹۳ | ۷-۷ خلاصه                               |

## فصل هشتم: استقرار مدل

۲۹۵

|     |                                             |
|-----|---------------------------------------------|
| ۲۹۶ | ۱-۸ استقرار ایستا                           |
| ۲۹۶ | ۲-۸ استقرار بویا روی دستگاه کاربر           |
| ۲۹۸ | ۳-۸ استقرار بویا در یک سرور                 |
| ۳۰۷ | ۴-۸ استراتژی های استقرار                    |
| ۳۰۹ | ۵-۸ استقرار، نسخه سازی و فراداده های خودکار |
| ۳۱۱ | ۶-۸ به روش هایی برای استقرار مدل            |
| ۳۲۰ | ۷-۸ خلاصه                                   |

## فصل نهم: استفاده، پایش و نگهداری مدل

۳۲۳

|     |                                         |
|-----|-----------------------------------------|
| ۳۲۴ | ۱-۹ خصوصیات استفاده از مدل در زمان اجرا |
| ۳۲۷ | ۲-۹ مدهای استفاده از مدل                |
| ۳۳۲ | ۳-۹ استفاده مدل در دنیای واقعی          |
| ۳۳۹ | ۴-۹ پایش مدل                            |

ذره خاکم و در کوی توام جای خوشست  
ترسم ای دوست که بادی ببرد ناگاهم

## پیشگفتار

در این سال‌ها، برای بسیاری از افراد یادگیری ماشین مترادف با هوش مصنوعی شده است. اگرچه یادگیری ماشین، به‌عنوان یک حوزه علمی، برای چندین دهه وجود داشته است، اما تعداد انگشت‌شماری از سازمان‌ها در جهان از پتانسیل آن به‌طور کامل استفاده کرده‌اند. با وجود در دسترس بودن کتابخانه‌های مدرن متن باز و بسته‌های نرم‌افزاری که توسط سازمان‌های پیشرو پشتیبانی می‌شود، اما بیشتر سازمان‌ها هنوز در تلاش برای استفاده از یادگیری ماشین برای حل مشکلات تجاری خود هستند. نکته اصلی همین است، منبع کاربردی مناسب در این حوزه زیاد نیست. با کمک یک کتاب معمولی یادگیری ماشین، می‌توانید با خانواده‌های اصلی الگوریتم‌ها، نحوه عملکرد آن‌ها و نحوه ایجاد مدل از داده‌ها با استفاده از آن الگوریتم‌ها آشنا شوید. این کتاب‌ها کمتر به جنبه‌های مهندسی اجرای پروژه‌های یادگیری ماشین توجه دارند.

خبر خوب اینکه کتاب حاضر از معدود کتاب‌هایی است که به‌خوبی جنبه‌های کاربردی پروژه‌های یادگیری ماشین را در خود دارد. فرض بر این است که خواننده این کتاب اصول اولیه یادگیری ماشین را می‌داند و با استفاده از یک زبان برنامه‌نویسی مورد علاقه یا یک کتابخانه یادگیری ماشین قادر به ساختن یک مدل هست. مخاطبان این کتاب تحلیلگران داده هستند که گرایش آن‌ها به سمت نقش مهندسی یادگیری ماشین است، مهندسان یادگیری ماشین که مایلند ساختار بیشتری را در کار خود داشته باشند و علاقه‌مندان به مهندسی یادگیری ماشین و همچنین معماران نرم‌افزاری که به‌طور معمول با مدل‌های ارائه شده توسط داده‌ها سروکار دارند.

مطالب کتاب در ده فصل گردآوری و تهیه شده است. در فصل اول کتاب، مسیر راهی (چرخه حیات پروژه‌های یادگیری ماشین) برای اجرای پروژه‌های یادگیری ماشین ارائه می‌شود که شامل نه مرحله است: تعریف هدف، جمع‌آوری و آماده‌سازی داده‌ها، مهندسی ویژگی، آموزش مدل، ارزیابی مدل، استقرار مدل، استفاده مدل، پایش مدل و نگهداری مدل. در فصل‌های بعدی هر یک از این مراحل به

تفصیل توضیح داده می‌شود. می‌توانید فصل‌ها را به هر ترتیب دلخواهی بخوانید، زیرا جنبه‌های متمایزی از چرخه حیات پروژه یادگیری ماشین را پوشش می‌دهند و وابستگی مستقیمی با یکدیگر ندارند. تمام تلاش خود را انجام داده‌ایم تا ترجمه و نگارش کتاب به گونه‌ای انجام شود که خوانندگان محترم آن بتوانند مفاهیم آن را به راحتی درک کنند. بدون شک ممکن است برای برخی از واژه‌های انگلیسی بتوان معادل‌های بهتری یافت. مسلم اینکه شاید گاهی تبدیل واژه‌ها آنچنان که باید و شاید انجام نشده است؛ اما در ادای جملات و بیان موضوع تلاش فراوان شده است تا خوانندگان گرمی با متنی مبهم و گیج‌کننده روبه‌رو نشوند.

در اینجا لازم می‌دانم از همه استادان و دانشجویان به‌خاطر راهنمایی‌های ارزشمندشان در حین آماده‌سازی این کتاب سپاسگزاری کنم. همچنین از مدیریت محترم انتشارات آتی‌نگر و دوست گرمی جناب رامین مولاناپور نیز به‌خاطر آماده‌سازی، چاپ و پخش این کتاب تشکر می‌کنم. رهین محبت بی‌دریغ خانواده‌ام هستم که با فراهم‌سازی محیطی مناسب مرا یاری کردند. اما با وجود همه سعی و تلاشی که در تمام مراحل آماده‌سازی این کتاب انجام گرفته است، یقین دارم که عاری از اشتباه نیست. در آخر ضمن سپاسگزاری از همه کسانی که مرا یاری داده‌اند و با پذیرش مسئولیت هرگونه کاستی احتمالی، امیدوارم که این اندک مفید افتد.

مهدی اسماعیلی، سیدمهدی وحیدی‌پور

شهریور ماه ۱۴۰۰