

# مهندسی ویژگی در یادگیری ماشین

مؤلف

پابلو دوبو

مترجم

ایوب ترکیان

نیاز دانش

سرشناسه	: دوبو، پابلو، ۱۹۷۶- م	Duboue, Pablo, 1976
عنوان و نام پدیدآور	: مهندسی ویژگی در یادگیری ماشین / مولف پابلو دوبو؛ مترجم ایوب ترکیان.	
مشخصات نشر	: تهران: نیاز دانش، ۱۳۹۹.	
مشخصات ظاهری	: ۲۴۸ ص: مصور، جدول، نمودار	
شابک	: 978-600-8906-96-4	
وضعیت فهرست نویسی	: فیبا	
یادداشت	: عنوان اصلی: .The art of feature engineering : essentials for machine learning,2020	
موضوع	: فراگیری ماشین	Machine learning
موضوع	: پایتون (زبان برنامه نویسی کامپیوتر)	Python (Computer program language)
شناسه افزوده	: ترکیان، ایوب، ۱۳۳۷ - ، مترجم	
رده بندی کنگره	: Q۳۲۵/۵	
رده بندی دیویی	: ۰۰۶/۳۱	
شماره کتابشناسی ملی	: ۷۵۳۰۴۴۹	
وضعیت رکورد	: فیبا	



نام کتاب	: مهندسی ویژگی در یادگیری ماشین
مؤلف	: پابلو دوبو
مترجم	: ایوب ترکیان
مدیر اجرایی - ناظر بر چاپ	: حمیدرضا محمد شیرازی - محمد شمس
ناشر	: نیاز دانش
صفحه آرا	: واحد تولید انتشارات نیازدانش
نوبت چاپ	: اول - ۱۴۰۰
شمارگان	: ۱۰۰ نسخه
قیمت	: ۵۰۰,۰۰۰ ریال

ISBN:978-600-8906-96-4

شابک: ۹۷۸-۶۰۰-۸۹۰۶-۹۶-۴

هرگونه چاپ و تکثیر (اعم از زیراکس، بازنویسی، ضبط کامپیوتری و تهیه‌ی CD) از محتویات این اثر بدون اجازه کتبی ناشر ممنوع است، متخلفان به موجب بند ۵ از ماده ۲ قانون حمایت از مؤلفان، مصنفان و هنرمندان تحت پیگرد قانونی قرار می‌گیرند.

کلیه حقوق این اثر برای ناشر محفوظ است.

تهران، میدان انقلاب، خیابان ۱۲ فروردین، تقاطع وحید نظری، پلاک ۲۵۵، طبقه ۱، واحد ۲

۰۹۱۲۷۰۷۳۹۳۵ - ۶۶۴۷۸۱۰۸ - ۶۶۴۷۸۱۰۶ - ۰۲۱

www.Niaze-Danesh.com

مشاوره جهت نشر: ۲۱۰۶۷۰۹ - ۰۹۱۲

## فهرست

۱۱	فصل اول: مقدمه
۱۴	۱-۱ مهندسی ویژگی
۱۷	۲-۱ ارزیابی
۱۸	۱-۲-۱ سنجها
۲۰	۲-۲-۱ اعتبارسنجی متقاطع
۲۱	۱-۲-۲-۱ تخمین خارج بخشی
۲۲	۳-۲-۱ فرابرازش
۲۳	۴-۲-۱ نفرین بُعدیت
۲۳	۳-۱ چرخه‌ها
۲۳	۱-۳-۱ چرخه یادگیری ماشین
۲۶	۲-۳-۱ چرخه مهندسی ویژگی
۲۸	۴-۱ تحلیل
۲۸	۱-۴-۱ تحلیل داده اکتشافی
۳۰	۲-۴-۱ تحلیل خطا
۳۱	۵-۱ فرایندهای دیگر
۳۲	۱-۵-۱ مدل سازی زمینه‌ای
۳۴	۲-۵-۱ ساخت ویژگی
۳۴	۱-۲-۵-۱ انواع ویژگی
۳۶	۲-۲-۵-۱ شابلون‌های ویژگی
۳۷	۶-۱ بحث
۳۹	فصل دوم: ترکیب ویژگی‌ها
۴۰	۱-۲-۱ نرمال سازی ویژگی‌ها
۴۳	۱-۱-۲ استاندارد سازی و هم‌بستگی زدایی
۴۵	۲-۱-۲ هموار سازی
۴۶	۱-۲-۱-۲ هموار سازی احتمال
۴۷	۳-۱-۲ وزن دهی ویژگی

۴۷	۱-۳-۱-۲- وزن دهی فرکانس سند معکوس
۴۸	۲-۳-۱-۲- کالیبراسیون دوربین
۴۸	۲-۲- گسسته‌سازی و سبببندی
۴۹	۱-۲-۲- گسسته‌سازی بدون نظارت
۴۹	۱-۱-۲-۲- سبببندی
۵۱	۲-۱-۲-۲- خوشه‌بندی
۵۲	۲-۲-۲- گسسته‌سازی با نظارت
۵۳	۱-۲-۲-۲- ChiMerge
۵۳	۲-۲-۲-۲- MDLP
۵۴	۳-۲-۲-۲- CAIM
۵۵	۳-۲- ویژگی‌های توصیفی
۵۵	۱-۳-۲- هیستوگرام‌ها
۵۷	۲-۳-۲- دیگر ویژگی‌های توصیفی
۵۸	۴-۲- برخورد با داده‌های پرت
۵۹	۱-۴-۲- آشکارسازی داده‌های پرت
۶۰	۵-۲- تکنیک‌های پیشرفته
۶۳	<b>فصل سوم: بسط ویژگی‌ها</b>
۶۴	۱-۳- ویژگی‌های قابل محاسبه
۷۱	۲-۳- جانپی
۷۴	۳-۳- تجزیه توابع پیچیده
۷۷	۴-۳- بسط ویژگی القا شده توسط کرنل
۸۳	<b>فصل چهارم: کاهش ویژگی‌ها</b>
۸۵	۱-۱-۴- سنجه‌ها
۸۵	۱-۱-۴- سنجه‌های استفاده ویژگی
۹۰	۲-۱-۴- سنجه‌های ویژگی‌های چندگانه
۹۲	۳-۱-۴- طبقه‌گرهای ویژگی تکی
۹۲	۴-۱-۴- زرورق
۹۳	۲-۱-۴- سرهم کردن مجموعه ویژگی: جستجو و فیلتر
۹۴	۱-۲-۴- مقتصدانه
۹۵	۲-۲-۴- معیار توقف

۹۶	..... ۳-۱-۴ تکنیک‌های پیشرفته
۹۷	..... ۱-۳-۱-۴ پایداری
۹۸	..... ۲-۳-۱-۴ لیست سیاه ویژگی‌ها
۹۸	..... ۲-۴ تنظیم و انتخاب ویژگی نهفته
۱۰۰	..... ۱-۲-۴ تنظیم $L_2$ : رگرسیون صخره
۱۰۰	..... ۲-۲-۴ تنظیم $L_1$ : LASSO
۱۰۱	..... ElasticNet ۱-۲-۲-۴
۱۰۲	..... ۳-۲-۴ الگوریتم‌های دیگر
۱۰۳	..... ۳-۴ کاهش بُعدیت
۱۰۴	..... ۱-۳-۴ هش کردن ویژگی‌ها
۱۰۵	..... ۲-۳-۴ افکنش راندومی
۱۰۶	..... SVD ۳-۳-۴
۱۰۷	..... ۴-۳-۴ تخصیص دیریکله نهفته
۱۰۸	..... ۵-۳-۴ خوشه‌بندی
۱۰۸	..... ۶-۳-۴ دیگر تکنیک‌های کاهش بُعدیت
۱۱۰	..... ۷-۳-۴ نهفته‌سازی
۱۱۲	..... ۱-۷-۳-۴ نهفته‌سازی فراگیر
۱۱۴	..... ۲-۷-۳-۴ نهفته‌سازی محلی: Word2Vec
۱۱۴	..... ۳-۷-۳-۴ نهفته‌سازی‌های دیگر: GloVe و ULMFiT
۱۱۵	..... فصل پنجم: موضوعات پیشرفته
۱۱۶	..... ۱-۵ بردارهای با طول متغیر
۱۱۶	..... ۱-۱-۵ مجموعه‌ها
۱۱۷	..... ۲-۱-۵ لیست‌ها
۱۱۹	..... ۳-۱-۵ درخت
۱۲۲	..... ۴-۱-۵ گراف
۱۲۲	..... ۵-۱-۵ سری زمانی
۱۲۳	..... ۱-۵-۱-۵ خودهم‌بستگی
۱۲۴	..... ۲-۵-۱-۵ روند، چرخه، و مؤلفه‌های فصلی
۱۲۵	..... ۳-۵-۱-۵ مانایی
۱۲۶	..... ۴-۵-۱-۵ مدل‌های سری زمانی: ARIMA

۱۲۷	۲-۵- مهندسی مورد پایه
۱۲۹	۳-۵- یادگیری عمیق و مهندسی ویژگی
۱۳۱	۳-۵-۱- RNN-ها
۱۳۳	۴-۵- مهندسی ویژگی اتوماتیک
۱۳۳	۵-۴-۱- یادگیری ویژگی
۱۳۳	۵-۴-۱-۱- شبکه‌های عصبی کانولوشن
۱۳۴	۵-۴-۱-۲- ابزار ویژگی
۱۳۶	۵-۴-۱-۳- برنامه‌سازی ژنتیک
۱۳۶	۵-۴-۲- مهندسی ویژگی بدون نظارت
۱۳۹	<b>فصل ششم: داده‌های گراف</b>
۱۴۱	مرور کلی فصل
۱۴۲	۶-۱- مجموعه داده شهرهای ویکی
۱۴۴	۶-۲- تحلیل داده اکتشافی (EDA)
۱۵۳	۶-۳- اولین مجموعه ویژگی
۱۵۴	۶-۳-۱- تحلیل خطا
۱۵۵	۶-۳-۱-۱- تمیز کردن ویژگی
۱۵۷	۶-۳-۱-۲- گسسته‌سازی هدف
۱۵۷	۶-۳-۱-۳- استفاده ویژگی
۱۵۸	۶-۳-۱-۴- درختان تصمیم برای تحلیل ویژگی
۱۵۹	۶-۴- دومین مجموعه ویژگی
۱۶۰	۶-۴-۱- پایداری ویژگی ایجاد مجموعه ویژگی محافظه‌کارانه
۱۶۲	۶-۵- مجموعه‌های ویژگی نهایی
۱۶۲	۶-۵-۱- پیگیری‌های بالقوه
۱۶۵	<b>فصل هفتم: داده‌های زمان‌مند</b>
۱۶۶	مرور فصل
۱۶۸	۷-۱- شهرهای ویکی: ویژگی‌های تاریخی
۱۷۰	۷-۱-۱- تحلیل داده اکتشافی
۱۷۱	۷-۲- ویژگی‌های با تأخیر زمانی
۱۷۲	۷-۲-۱- جانهای داده‌های زمان‌مند
۱۷۳	۷-۲-۲- اولین ویژگی‌سازی: تأخیر-۲ جانهای شده

۱۷۳	..... فرض مارکوف..... ۱-۲-۲-۷
۱۷۴	..... ویژگی‌های دیفرانسیل..... ۲-۲-۲-۷
۱۷۴	..... تحلیل خطا..... ۳-۲-۷
۱۷۴	..... پنجره‌های متحرک..... ۳-۷
۱۷۶	..... ۱-۳-۷ دومین ویژگی‌سازی: میانگین متحرک تکی.....
۱۷۶	..... ۴-۷ ویژگی‌سازی سوم: EMA.....
۱۷۶	..... ۵-۷ داده‌های تاریخی به عنوان بسط داده‌ها.....
۱۷۷	..... ۱-۵-۷ ویژگی‌سازی چهارم: داده‌های بسط یافته.....
۱۷۷	..... ۲-۵-۷ بحث.....
۱۷۸	..... ۶-۷ سری زمانی.....
۱۷۸	..... ۱-۶-۷ مجموعه داده کشورهای ویکی.....
۱۷۹	..... ۲-۶-۷ تحلیل داده اکتشافی.....
۱۸۳	..... ۳-۶-۷ اولین ویژگی‌سازی: ویژگی‌های NoTS.....
۱۸۳	..... ۴-۶-۷ دومین ویژگی‌سازی: TS به عنوان ویژگی.....
۱۸۴	..... ۵-۶-۷ استفاده از پیش‌بینی مدل به عنوان ویژگی.....
۱۸۵	..... ۶-۶-۷ بحث.....
۱۸۶	..... ۷-۷ یادگیری بیشتر.....
۱۸۹	..... فصل هشتم: داده‌های متنی.....
۱۹۱	..... مرور فصل.....
۱۹۲	..... ۱-۸ شهرهای ویکی: متن.....
۱۹۳	..... ۲-۸ تحلیل داده اکتشافی.....
۱۹۷	..... ۳-۸ فقط توکن‌های عددی.....
۱۹۷	..... ۱-۳-۸ انواع واژه در مقابل توکن‌های واژه.....
۱۹۷	..... ۲-۳-۸ توکن‌سازی: مبانی.....
۱۹۹	..... ۳-۳-۸ اولین ویژگی‌سازی.....
۱۹۹	..... ۴-۸ کیف‌واژه.....
۲۰۰	..... ۱-۴-۸ توکن‌سازی.....
۲۰۲	..... ۲-۴-۸ دومین ویژگی‌سازی.....
۲۰۲	..... ۱-۲-۴-۸ تحلیل خطا.....
۲۰۳	..... ۵-۸ واژه‌های توقف و ویژگی‌های مورفولوژیکی.....

- ۲۰۴..... ۱-۵-۸ واژه‌های توقف
- ۲۰۴..... ۲-۵-۸ توکن‌سازی: ریشه‌یابی
- ۲۰۵..... ۳-۵-۸ سومین ویژگی‌سازی
- ۲۰۶..... ۱-۳-۵-۸ تحلیل خطا
- ۲۰۶..... ۶-۸ ویژگی‌ها در بافت
- ۲۰۶..... ۱-۶-۸ بایگرام‌ها
- ۲۰۷..... ۲-۶-۸ چهارمین ویژگی‌سازی
- ۲۰۷..... ۷-۸ بایگرام‌های گذر و هش کردن ویژگی
- ۲۰۸..... ۱-۷-۸ بایگرام گذر
- ۲۰۸..... ۲-۷-۸ پنجمین ویژگی‌سازی
- ۲۰۹..... ۸-۸ کاهش بُعدیت و نهفته‌سازی
- ۲۰۹..... ۱-۸-۸ نهفته‌سازی‌ها
- ۲۱۰..... ۲-۸-۸ وزن‌دهی ویژگی: TF-IDF
- ۲۱۱..... ۳-۸-۸ ششمین ویژگی‌سازی
- ۲۱۱..... ۹-۸ اشارات پایانی
- ۲۱۳..... ۱-۹-۸ بسط محتوا
- ۲۱۴..... ۲-۹-۸ ساختار در متن
- ۲۱۵..... فصل نهم: داده‌های تصویری**
- ۲۱۶..... مرور فصل
- ۲۱۷..... ۱-۹ شهرهای ویکی: تصاویر ماهواره‌ای
- ۲۱۸..... ۲-۹ تحلیل داده اکتشافی
- ۲۱۹..... ۳-۹ پیکسل به عنوان ویژگی
- ۲۲۰..... ۱-۳-۹ اولین ویژگی‌سازی
- ۲۲۰..... ۱-۱-۳-۹ تحلیل خطا
- ۲۲۲..... ۲-۳-۹ ویژگی‌های قابل محاسبه: تیرگی گوسی
- ۲۲۳..... ۳-۳-۹ سفیدسازی
- ۲۲۵..... ۴-۳-۹ تحلیل خطا روی تغییرپذیری‌ها
- ۲۲۶..... ۴-۹ بسط مجموعه داده اتوماتیک
- ۲۲۶..... ۱-۴-۹ تبدیل‌های آفین
- ۲۲۶..... ۲-۴-۹ دومین ویژگی‌سازی



- ۲۲۷..... ۹-۴-۱- تحلیل خطا
- ۲۲۷..... ۹-۵- ویژگی‌های توصیفی: هیستوگرام‌ها
- ۲۲۷..... ۹-۵-۱- سومین ویژگی‌سازی
- ۲۲۸..... ۹-۵-۱- تحلیل خطا
- ۲۲۸..... ۹-۶- آشکارسازهای ویژگی محلی: گوشه‌ها
- ۲۲۹..... ۹-۶-۱- آشکارساز گوشه هریس
- ۲۲۹..... ۹-۶-۲- چهارمین ویژگی‌سازی
- ۲۳۰..... ۹-۶-۲-۱- تحلیل خطا
- ۲۳۱..... ۹-۷- کاهش بُعدیت: HOG-ها
- ۲۳۲..... ۹-۷-۱- پنجمین ویژگی‌سازی
- ۲۳۲..... ۹-۸- اشارات پایانی
- ۲۳۳..... ۹-۸-۱- موضوعات دیگر در بینایی کامپیوتر
- ۲۳۵..... فصل دهم: زمینه‌های دیگر**
- ۲۳۶..... ۱۰-۱- ویدئو
- ۲۳۷..... ۱۰-۱-۱- داده‌ها: نمایش صفحه
- ۲۳۷..... ۱۰-۱-۲- آشکارسازی قاب کلیدی
- ۲۳۹..... ۱۰-۱-۳- ردیابی لکه‌ها: شیفت میانگین
- ۲۳۹..... ۱۰-۱-۳-۱- پس‌افکنش هیستوگرام
- ۲۳۹..... ۱۰-۱-۳-۲- شیفت میانگین
- ۲۴۰..... ۱۰-۱-۴- یادگیری بیشتر
- ۲۴۱..... ۱۰-۲- ویژگی‌های جغرافیایی
- ۲۴۱..... ۱۰-۲-۱- داده‌ها: مهاجرات حیوان
- ۲۴۲..... ۱۰-۲-۱- فاصله شعاعی به علامات راهنما
- ۲۴۳..... ۱۰-۲-۱-۲- یادگیری بیشتر
- ۲۴۳..... ۱۰-۳- ترجیحات
- ۲۴۴..... ۱۰-۳-۱- داده‌ها: کرنل تعهدهای لینوکس
- ۲۴۵..... ۱۰-۳-۲- جانهی ترجیحات
- ۲۴۶..... ۱۰-۳-۳- یادگیری بیشتر

