

به نام خدا



# علم داد از تئوری تا عمل

مؤلفان

محمد جواد جعفری

کارشناسی ارشد مهندسی فناوری اطلاعات

پرمان محمد علیزاده

کارشناسی معماری کامپیوتر

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.

## علم داده از تئوری تا عمل

مؤلفا: محمد جواد جعفری

رمان محمد علیزاده

ناشر: مؤسسه فرهنگی هنری دیباگران تهران

حروفچینی: صفحه آبی: شبنم هاشم زاده

طرح روی جلد: آریوش خرابایی

چاپ: دانشجو

نوبت چاپ: اول

تاریخ نشر: ۱۳۹۶

تیراژ: ۱۰۰ جلد

قیمت: ۴۰۰۰۰ ریال

شابک: ۹۷۸-۶۰۰-۱۲۴-۵۸۰-۰

نشانی واحد فروش: تهران، میدان انقلاب،

خ کارگر جنوبی، روبروی پاساژ مهستان،

پلاک ۱۲۵۱

تلفن: ۰۲۲۰۸۵۱۱۱-۶۶۴۱۰۰۴۶ کد پستی: ۱۳۱۴۹۸۳۱۸۵

پست الکترونیکی: [bookmarket@mftmail.com](mailto:bookmarket@mftmail.com)

فروش اینترنتی

[www.mftshop.com](http://www.mftshop.com)

[www.mftbook.ir](http://www.mftbook.ir)

نشانی اینستاگرام: [Dibagaran\\_publishing](https://www.instagram.com/Dibagaran_publishing)

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

سرشناسه: جعفری، محمد جواد، ۱۳۶۷-

عنوان و نام پدید آور: علم داده از تئوری تا عمل / مؤلفان:  
محمد جواد جعفری، پرمان محمد علیزاده

مشخصات نشر: تهران- دیباگران تهران- ۱۳۹۶

مشخصات ظاهری: ۳۷۰ ص. مصور.

شابک: ۹۷۸-۶۰۰-۱۲۴-۵۸۰-۰

وضعیت فهرست نویسی: فیبا

موضوع: داده کاوی

موضوع: Data mining

شناسه افزوده: محمد علیزاده، پرمان، ۱۳۷۴-

ردیف پند: ۱۳۹۶ ج ۶ QA ۷۶/۹/د ۲

ردیف پند دیوبی: ۰۰۶/۳۱۲

شمار کتابخانه: ۴۶۶۳۸۷۷

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| بخش اول.....                                                            | ۲۱ |
| شروع کار با علم داده .....                                              | ۲۱ |
| فصل اول.....                                                            | ۲۳ |
| نگاهی اجمالی بر علم داده‌ها .....                                       | ۲۳ |
| بررسی راه‌کارهای پیشنهادی علم داده .....                                | ۲۴ |
| ایجاد یک تیم درون سازمانی .....                                         | ۲۴ |
| برون‌سپاری کارها به مشاوران خصوصی علم داده .....                        | ۲۵ |
| استفاده از راه‌حل‌های ابری محو .....                                    | ۲۶ |
| مزایای محسوس علم داده .....                                             | ۲۶ |
| فصل دوم.....                                                            | ۲۷ |
| بررسی زیرساخت‌های مهندسی داده .....                                     | ۲۷ |
| تعریف کلان داده با ۴ ویژگی اصلی آن .....                                | ۲۸ |
| تنوع داده‌ها .....                                                      | ۳۰ |
| منابع کلان داده کدامند؟ .....                                           | ۳۱ |
| تفاوت بین مهندسی داده و علم داده .....                                  | ۳۱ |
| تعریف علم داده .....                                                    | ۳۲ |
| تعریف مهندسی داده .....                                                 | ۳۲ |
| مقایسه متخصصین داده و مهندسان داده .....                                | ۳۳ |
| محدود کردن داده‌ها با استفاده از <i>hadoop</i> و <i>mapreduce</i> ..... | ۳۴ |
| نگاهی عمیق‌تر به <i>mapreduce</i> .....                                 | ۳۴ |

|    |                                                          |
|----|----------------------------------------------------------|
| ۳۵ | نگاشت داده‌ها                                            |
| ۳۵ | کاهش داده                                                |
| ۳۶ | Hadoop چیست؟                                             |
| ۳۹ | معرفی چارچوب‌های پردازش بلادرنگ                          |
| ۴۰ | معرفی پلتفرم‌های پردازش موازی عظیم Mpp                   |
| ۴۱ | معرفی پایگاه داده‌های NoSQL                              |
| ۴۲ | درسی یک نمونه عملی از مهندسی داده                        |
| ۴۲ | شناسایی چالش‌های مربوط به کسب‌وکار در این زمینه          |
| ۴۳ | ملاحظات مربوط به کسب‌وکار با استفاده از مهندسی داده      |
| ۴۳ | چه مزایایی برای این شرکت حاصل شد؟                        |
| ۴۵ | فصل سوم                                                  |
| ۴۵ | کاربرد علم داده در صنعت کسب‌وکار                         |
| ۴۶ | ترکیب بینش‌های حاصل‌شده از داده‌ها در فرآیندهای کسب‌وکار |
| ۴۶ | مزایای علم داده‌ی کسب‌وکار محور                          |
| ۴۷ | شناسایی چالش‌های رایج در تجزیه و تحلیل                   |
| ۴۷ | گردآوری داده‌های خام و تبدیل آن‌ها به بینش عملی          |
| ۴۹ | اقدام بر اساس بینش‌های حاصل‌شده                          |
| ۵۰ | تفاوت هوش تجاری و علم داده                               |
| ۵۱ | تعریف هوش تجاری                                          |
| ۵۶ | نگاهی به انواع داده‌ی مورد استفاده در هوش تجاری          |
| ۵۶ | بررسی تکنولوژی و مهارت‌هایی که در هوش تجاری مفید هستند   |
| ۵۸ | فرایند ETL                                               |
| ۵۸ | تعریف علم داده کسب‌وکار محور                             |
| ۵۹ | بررسی انواع داده‌های مفید در علم داده‌ی کسب‌وکار محور    |

تفاوت‌های اصلی هوش تجاری و علم داده کسب و کار محور ۶۱

علم داده در تجارت (داستان موفقیت یک تجارت مبتنی بر داده) ۶۳

## بخش دوم ۶۵

استفاده از علم داده در راستای استخراج بینش از داده‌ها ۶۵

## فصل چهارم ۶۷

معرفی آمار و احتمالات ۶۷

معرفی مفاهیم اساسی احتمال ۶۸

بررسی رابطه بین احتمال و آمار استنباطی ۶۸

آشنایی با برخی تمایزهای احتمال رایج ۷۰

معرفی رگرسیون خطی ۷۲

آشنایی با مدل‌های داده رگرسیونی خطی ۷۲

آموزش ساخت خط رگرسیون با اندازه تناسب ۷۴

روش‌هایی ساده جهت محاسبه رگرسیون حداقل مربعات معمولی شبیه‌سازی ۷۵

چگونه این عدد تصادفی به رنگ چشم مرتبط است؟ ۷۹

استفاده از شبیه‌سازی برای ارزیابی خواص یک آمارگر ۸۱

استفاده از شبیه‌سازی مونت کارلو برای ارزیابی خواص یک برآوردگر ۸۲

معرفی تجزیه و تحلیل سری‌های زمانی ۸۵

درک الگوهای سری زمانی ۸۵

مدل‌سازی داده‌های سری زمانی تک متغیره ۸۶

## فصل پنجم ۸۹

خوشه‌بندی و طبقه‌بندی ۸۹

معرفی اصول خوشه‌بندی و طبقه‌بندی ۹۰

درک الگوریتم‌های خوشه‌بندی ۹۱

|     |                                                                        |
|-----|------------------------------------------------------------------------|
| ۹۶  | خوشه‌بندی با استفاده از الگوریتم $k$ -means                            |
| ۹۷  | روش KDE                                                                |
| ۹۸  | خوشه‌بندی با الگوریتم‌های سلسله مراتبی و محلی                          |
| ۱۰۰ | طبقه‌بندی داده‌ها با درخت تصمیم و الگوریتم‌های جنگل تصادفی             |
| ۱۰۳ | فصل ششم                                                                |
| ۱۰۳ | خوشه‌بندی و طبقه‌بندی                                                  |
| ۱۰۳ | با استفاده از الگوریتم نزدیک‌ترین همسایه                               |
| ۱۰۴ | مفهوم بخشی به داده‌ها با استفاده از تجزیه و تحلیل نزدیک‌ترین همسایه    |
| ۱۰۵ | درک اهمیت خردی و طبقه‌بندی                                             |
| ۱۰۶ | طبقه‌بندی داده‌ها با استفاده از الگوریتم‌های میانگین نزدیک‌ترین همسایه |
| ۱۰۷ | مقایسه شباهت‌های میانگین با تحلیلگر کسب‌وکار $Stu$                     |
| ۱۰۹ | طبقه‌بندی با الگوریتم‌های $k$ آمیخته نزدیک‌ترین همسایه                 |
| ۱۱۱ | چه موقع باید از الگوریتم $KNN$ استفاده کنیم؟                           |
| ۱۱۲ | کاربردهای رایج الگوریتم $KNN$                                          |
| ۱۱۲ | استفاده از نزدیک‌ترین فاصله برای استخراج مفاهیم از مجموعه‌ی نقاط       |
| ۱۱۳ | حل مسائل دنیای واقعی به وسیله‌ی الگوریتم‌های نزدیک‌ترین همسایه         |
| ۱۱۶ | استفاده از روش‌های عددی در علم داده‌ها                                 |
| ۱۱۷ | بسط چندجمله‌ای تیلور                                                   |
| ۱۱۹ | دوبخشی کردن توابع با استفاده از الگوریتم جستجوی دوبخشی                 |
| ۱۲۰ | مدل‌سازی ریاضیاتی با زنجیره‌های مارکوف و روش‌های تصادفی                |
| ۱۲۳ | فصل هفتم                                                               |
| ۱۲۳ | مدل‌سازی داده‌های فضایی با آمار                                        |
| ۱۲۴ | تولید سطح قابل پیش‌بینی از نقاط داده‌های فضایی                         |
| ۱۲۵ | آشنایی با $(x, y, z)$ مدل داده‌های فضایی                               |

|     |                                                                                        |
|-----|----------------------------------------------------------------------------------------|
| ۱۲۶ | معرفی <i>kriging</i> .....                                                             |
| ۱۲۷ | <i>Krige</i> کردن به منظور درون‌یابی <i>kriging</i> خودکار .....                       |
| ۱۲۷ | انتخاب و استفاده از مدل‌های مناسب برای درون‌یابی <i>kriging</i> صریح و تعریف‌شده ..... |
| ۱۲۷ | میانگیری با مدل‌های <i>variogram</i> صریح و تعریف‌شده .....                            |
| ۱۲۸ | تخمین <i>variogram</i> .....                                                           |
| ۱۲۹ | نگاهی دقیق به روش <i>kriging rabbit hole</i> .....                                     |
| ۱۲۹ | بررسی روش‌های تخمین سطح <i>kriging</i> های معمولی .....                                |
| ۱۳۰ | بررسی روش‌های تخمین سطح رگرسیون <i>kriging</i> .....                                   |
| ۱۳۰ | بررسی تخمین زن‌های سطح <i>kriging</i> بلوکی .....                                      |
| ۱۳۲ | بررسی تخمین زن‌های سطح <i>cokriging</i> .....                                          |
| ۱۳۳ | انتخاب بهترین روش تخمین در <i>kriging</i> .....                                        |
| ۱۳۵ | تحلیل باقی‌مانده‌ها برای تعیین بهترین مدل مناسب .....                                  |
| ۱۳۸ | شناخت گزینه‌های موجود در <i>kriging</i> .....                                          |
| ۱۳۸ | استفاده از تحلیل سطح روند بر داده‌های نمایشی .....                                     |
| ۱۴۱ | بخش سوم .....                                                                          |
| ۱۴۱ | ایجاد بصری سازی‌های معنادار از داده‌ها .....                                           |
| ۱۴۳ | فصل هشتم .....                                                                         |
| ۱۴۳ | به‌کارگیری اصول طراحی بصری‌سازی داده .....                                             |
| ۱۴۴ | شناخت انواع بصری‌سازی .....                                                            |
| ۱۴۴ | توصیف داستانی داده‌ها برای تصمیم‌گیرندگان سازمانی: .....                               |
| ۱۴۵ | نمایش داده‌ها برای تحلیل‌گران .....                                                    |
| ۱۴۵ | طراحی تابلوی نمایش داده برای فعالان این زمینه .....                                    |
| ۱۴۵ | تمرکز بر مخاطبان .....                                                                 |
| ۱۴۸ | انتخاب مناسب‌ترین حالت طراحی .....                                                     |

|     |                                                         |
|-----|---------------------------------------------------------|
| ۱۴۸ | ..... استفاده از بصریسازی برای درک دقیق محاسباتی        |
| ۱۴۹ | ..... استفاده از طراحی برای ایجاد شور و اشتیاق در مخاطب |
| ۱۵۱ | ..... استفاده از داده‌ها برای ایجاد متن                 |
| ۱۵۱ | ..... ایجاد متن با استفاده از حاشیه‌نویسی               |
| ۱۵۳ | ..... ایجاد متن با استفاده از عناصر گرافیکی             |
| ۱۵۳ | ..... چه زمانی باید از حالت متقاعدکننده استفاده کنیم؟   |
| ۱۵۳ | ..... انتخاب مناسب‌ترین نوع داده گرافیکی                |
| ۱۵۴ | ..... بررسی نمودار استاندارد گرافیکی                    |
| ۱۵۷ | ..... بررسی گرافیک متنی                                 |
| ۱۶۱ | ..... بررسی طرح‌های تباری                               |
| ۱۶۳ | ..... بررسی ساختارهای ریولوشن                           |
| ۱۶۵ | ..... بررسی طرح‌ها و نقشه‌ها: فضا                       |
| ۱۶۷ | ..... انتخاب گرافیک داده شما                            |
| ۱۶۸ | ..... تعیین دامنه سؤالات                                |
| ۱۶۸ | ..... در نظر گرفتن رسانه و کاربر                        |
| ۱۶۸ | ..... نگاهی مجدد به گام نهایی بیندازید                  |
| ۱۶۹ | ..... فصل نهم                                           |
| ۱۶۹ | ..... استفاده از <code>DOM</code> برای بصریسازی داده    |
| ۱۷۰ | ..... معرفی کتابخانه <code>DOM</code>                   |
| ۱۷۱ | ..... چه زمانی باید از <code>DOM</code> استفاده کنیم؟   |
| ۱۷۲ | ..... شروع به کار در <code>DOM</code>                   |
| ۱۷۲ | ..... بررسی <code>HTML</code> و <code>DOM</code>        |
| ۱۷۳ | ..... بررسی <code>JavaScript</code> و <code>SVG</code>  |
| ۱۷۵ | ..... بررسی سرورهای وب و <code>PHP</code>               |



|     |                                                                  |
|-----|------------------------------------------------------------------|
| ۱۷۶ | شناخت مفاهیم پیشرفته تر و شیوه عملکرد <i>LeafletJS</i>           |
| ۱۸۰ | بررسی ساختار زنجیره ای                                           |
| ۱۸۱ | بررسی مقیاس ها                                                   |
| ۱۸۲ | مروری بر انتقال ها و تعاملات                                     |
| ۱۸۵ | فصل دهم                                                          |
| ۱۸۵ | اپلیکیشن مبتنی بر وب برای طراحی بصری سازی                        |
| ۱۸۷ | کار با تحلیلگر واتسون <i>IBM</i>                                 |
| ۱۸۸ | بصری سازی و کار تیمی با استفاده از <i>PLOTLY</i>                 |
| ۱۹۱ | بصری سازی داده های فضایی با استفاده از ابزارهای جغرافیایی آنلاین |
| ۱۹۲ | ایجاد نقشه های خیره کننده با <i>Open Heat Map</i>                |
| ۱۹۳ | ایجاد نقشه و تبدیل داده های فضایی با <i>CartoDB</i>              |
| ۱۹۵ | بصری سازی با ابزار متن باز: ابزارهای بصری سازی داده مبتنی بر وب  |
| ۱۹۵ | ایجاد داده های گرافیکی زیبا با <i>Google fusion</i>              |
| ۱۹۶ | استفاده از <i>iCharts</i> برای بصری سازی داده مبتنی بر وب        |
| ۱۹۷ | استفاده از <i>RAW</i> برای بصری سازی داده مبتنی بر وب            |
| ۱۹۹ | چه زمانی باید از اینفوگرافیک استفاده کنیم؟                       |
| ۱۹۹ | ساختن اینفوگرافیک های جذاب با استفاده از <i>fogr.am</i>          |
| ۲۰۱ | مهاجرت درون ایالتی                                               |
| ۲۰۱ | ایجاد اینفوگرافیک های جذاب با استفاده از <i>PiktoChart</i>       |
| ۲۰۳ | فصل یازدهم                                                       |
| ۲۰۳ | بررسی بهترین روش ها در طراحی داشبورد                             |
| ۲۰۴ | تمرکز بر مخاطبین                                                 |
| ۲۰۵ | طرح کلی                                                          |
| ۲۰۶ | طراحی جزئیات                                                     |

|     |                                                    |
|-----|----------------------------------------------------|
| ۲۰۹ | فصل دوازدهم                                        |
| ۲۰۹ | ساخت نقشه با استفاده از داده‌های فضایی             |
| ۲۱۰ | آشنایی با اساس GIS                                 |
| ۲۱۲ | فرمت‌های فایل‌های GIS                              |
| ۲۱۵ | درک مفهوم نقشه و سیستم‌های مختصات                  |
| ۲۱۸ | بافر و توابع مجاورت                                |
| ۲۱۹ | ارائه داده از تحلیل پوشش لایه‌ای                   |
| ۲۲۰ | مجموعه‌بندی مجدد داده‌های فضایی                    |
| ۲۲۱ | عملیات ساده اعمال شده بر ویژگی‌های دارای هم‌پوشانی |
| ۲۲۱ | کار با GIS از منظر                                 |
| ۲۲۱ | آشنایی با رابط GIS                                 |
| ۲۲۲ | اضافه کردن یک لایه بردار در QGIS                   |
| ۲۲۳ | نمایش داده در QGIS                                 |
| ۲۲۹ | بخش چهارم                                          |
| ۲۲۹ | محاسبات در علم داده                                |
| ۲۳۱ | فصل سیزدهم                                         |
| ۲۳۱ | استفاده از PYTHON برای علم داده                    |
| ۲۳۲ | آشنایی با مفاهیم اساسی در PYTHON                   |
| ۲۳۳ | معرفی انواع داده PYTHON                            |
| ۲۳۴ | اعداد در PYTHON                                    |
| ۲۳۴ | رشته‌ها در PYTHON                                  |
| ۲۳۴ | لیست‌ها در PYTHON                                  |
| ۲۳۵ | تاپل‌ها در PYTHON                                  |
| ۲۳۵ | مجموعه‌ها در PYTHON                                |

|     |                                                                           |
|-----|---------------------------------------------------------------------------|
| ۲۳۶ | استفاده حلقه‌ها در <i>PYTHON</i> .....                                    |
| ۲۳۷ | آشنایی با توابع و کلاس‌ها .....                                           |
| ۲۳۷ | معرفی توابع .....                                                         |
| ۲۳۹ | مزیت‌های استفاده از کلاس‌ها .....                                         |
| ۲۴۱ | معرفی کتابخانه‌ی <i>Numpy</i> .....                                       |
| ۲۴۳ | آشنایی بیشتر با کتابخانه <i>SciPy</i> .....                               |
| ۲۴۴ | استفاده از <i>MatPlotLib</i> برای بصرسازی داده .....                      |
| ۲۴۵ | استفاده از <i>PYTHON</i> برای تحلیل داده - یک نمونه عملی .....            |
| ۲۴۶ | نصب <i>PYTHON</i> بر روی سیستم‌عامل‌های <i>Mac</i> و <i>Windows</i> ..... |
| ۲۴۷ | بارگذاری فایل‌های <i>SV</i> .....                                         |
| ۲۴۸ | محاسبه میانگین و وزن .....                                                |
| ۲۵۰ | رسم خطوط روند .....                                                       |
| ۲۵۲ | فصل چهاردهم .....                                                         |
| ۲۵۳ | استفاده از زبان <i>R</i> در علم داده .....                                |
| ۲۵۴ | تسلط بر لغات مربوط به زبان <i>R</i> .....                                 |
| ۲۵۶ | محاسبات مربوط به علم داده .....                                           |
| ۲۵۷ | آشنایی بیشتر با توابع و عملگرها .....                                     |
| ۲۶۰ | پیمایش در <i>R</i> .....                                                  |
| ۲۶۱ | آشنایی با عملکرد اشیاء .....                                              |
| ۲۶۴ | نگاهی بر پکیج‌های <i>R</i> .....                                          |
| ۲۶۴ | نگاهی بر پکیج‌های مشهور تجزیه و تحلیل آماری .....                         |
| ۲۶۵ | بصری سازی، نگاشت و نمایش نموداری در <i>R</i> .....                        |
| ۲۶۵ | بصری سازی آماری <i>R</i> با استفاده از <i>ggplot۲</i> .....               |

|     |                                                                                  |
|-----|----------------------------------------------------------------------------------|
| ۲۶۶ | تجزیه و تحلیل شبکه‌ها با استفاده از <i>igraph</i> و <i>statnet</i>               |
| ۲۶۷ | نقشه‌برداری و تجزیه و تحلیل الگوی نقطه‌ای فضایی با به کارگیری <i>spatstat</i>    |
| ۲۶۹ | فصل پانزدهم                                                                      |
| ۲۶۹ | استفاده از <i>SQL</i> در علم داده                                                |
| ۲۷۰ | آغاز کار با <i>SQL</i>                                                           |
| ۲۷۰ | بررسی پایگاه داده‌های رابطه‌ای و <i>SQL</i>                                      |
| ۲۷۴ | آشنایی با مفهوم طراحی پایگاه داده                                                |
| ۲۷۴ | تعریف انواع داده                                                                 |
| ۲۷۵ | توجه به قواعد امنیت                                                              |
| ۲۷۵ | نرمال سازی پایگاه داده                                                           |
| ۲۷۷ | استفاده از <i>SQL</i> و سریع آ در علم داده                                       |
| ۲۷۷ | یکپارچه سازی <i>SQL</i> ، <i>R</i> ، <i>HCV</i> و اکسل در استراتژی علم داده      |
| ۲۷۷ | استفاده از توابع <i>SQL</i> در علم داده                                          |
| ۲۸۲ | متن کاوی با استفاده از <i>SQL</i>                                                |
| ۲۸۳ | فصل شانزدهم                                                                      |
| ۲۸۳ | نرم افزارهای متناسب با علم داده                                                  |
| ۲۸۴ | مزیت‌های تسهیل کننده‌ی اکسل                                                      |
| ۲۸۵ | استفاده از اکسل برای درک سریع داده‌ها                                            |
| ۲۸۶ | استفاده از فیلتر در اکسل                                                         |
| ۲۸۶ | فرمت بندی شرطی برای نشانه گذاری داده‌های خارج از محدوده و روندها                 |
| ۲۸۷ | رسم نمودار در اکسل برای مشاهده داده‌های خارج از محدوده و روندها                  |
| ۲۸۹ | فرمت بندی مجدد و خلاصه سازی با استفاده از جداول محوری                            |
| ۲۹۰ | خودکارسازی وظایف اکسل با استفاده از ماکرو (دستورات به کار رفته در <i>Excel</i> ) |
| ۲۹۲ | استفاده از <i>KNIME</i> برای تحلیل پیشرفته داده‌ها                               |

اعمال مهارت‌های حوزه‌های خاص ..... ۲۹۷

برای حل مسائل حقیقی با استفاده از علم داده ..... ۲۹۷

فصل هفدهم ..... ۲۹۹

بررسی تأثیر علم داده در خبرنگاری ..... ۲۹۹

چه کسی داده‌ها را تولید میکند؟ ..... ۳۰۱

نگاهی دقیق‌تر به اهمیت داستان خبری ..... ۳۰۲

چرا موضوعات شما برای مخاطبان اهمیت دارد؟ ..... ۳۰۲

بهترین زمان برای انتشار مطالب انتخاب کنید ..... ۳۰۵

چگونه باید داستان‌های هفت در داده‌های خود را دریابید؟ ..... ۳۰۶

چگونه باید داستان‌های محرک خود را ارائه دهید؟ ..... ۳۰۷

جمع‌آوری داده ..... ۳۰۷

استخراج داده‌ها برای داستان‌های خبر ..... ۳۰۸

بررسی روندهای غیر معمول و داده‌های خارج از محدوده ..... ۳۰۸

تأکید کردن بر داستان از طریق بصرسازی ..... ۳۱۱

فصل هجدهم ..... ۳۱۵

نگاهی به علم داده محیطی ..... ۳۱۵

مدل‌سازی تعامل انسان و محیط با استفاده از هوش محیطی ..... ۳۱۶

نگاهی به انواع مشکلات حل‌شده ..... ۳۱۶

تعریف هوش محیطی ..... ۳۱۷

شناسایی سازمان‌های بزرگ که در حوزه هوش محیطی کار میکنند ..... ۳۱۸

تأثیرگذاری از طریق هوش محیطی ..... ۳۱۹

مدل‌سازی منابع طبیعی خام ..... ۳۲۱

بررسی مدل‌سازی منابع طبیعی ..... ۳۲۱

|          |                                                               |
|----------|---------------------------------------------------------------|
| ۳۲۲..... | مدل سازی منابع طبیعی برای حل مشکلات زیست محیطی .....          |
| ۳۲۴..... | توصیف نقش علم داده در مدل سازی آمار فضایی .....               |
| ۳۲۵..... | استفاده از آمار فضایی برای رسیدگی به مسائل زیست محیطی .....   |
| ۳۲۷..... | فصل نوزدهم .....                                              |
| ۳۲۷..... | استفاده از علم داده برای رشد تجارت الکترونیکی .....           |
| ۳۳۰..... | ایجاد درکی شفاف از داده ها به منظور رشد تجارت الکترونیک ..... |
| ۳۳۱..... | بهسازی سیستم های تجارت الکترونیک .....                        |
| ۳۳۲..... | بررسی جنبه های تحلیلی .....                                   |
| ۳۳۳..... | ارزیابی اثربخشی های پیرامون تحلیل وب .....                    |
| ۳۳۴..... | دسترسی به تجزیه ها به منظور یافتن دستاورد .....               |
| ۳۳۵..... | استفاده از تحلیل برای فعالیت .....                            |
| ۳۳۶..... | بررسی تحلیل ها برای حفظ کاربر .....                           |
| ۳۳۷..... | بررسی و آزمایش استراتژی .....                                 |
| ۳۳۷..... | جمع بندی انواع روش های آزمایش در فرایند رشد .....             |
| ۳۳۹..... | تست دستیابی .....                                             |
| ۳۳۹..... | آزمایش فعالیت .....                                           |
| ۳۴۰..... | بخش بندی و هدف گذاری برای رسیدن به موفقیت .....               |
| ۳۴۱..... | تقسیم بندی برای رشد سریع و راحت تر تجارت الکترونیک .....      |
| ۳۴۲..... | پیدا کردن مخاطبان .....                                       |
| ۳۴۲..... | بهینه سازی کانال های شبکه های اجتماعی .....                   |
| ۳۴۳..... | بخش بندی و هدفمند سازی برای حفظ کاربران .....                 |
| ۳۴۳..... | بخش بندی و هدف قرار دادن درآمدزایی .....                      |
| ۳۴۵..... | فصل بیستم .....                                               |
| ۳۴۵..... | استفاده از علم داده برای توصیف و پیش بینی اعمال مجرمانه ..... |

|     |                                                   |
|-----|---------------------------------------------------|
| ۳۴۶ | تجزیه و تحلیل زمانی برای پیشگیری از جرائم         |
| ۳۴۷ | پیش‌بینی و نظارت بر جرائم فضایی                   |
| ۳۴۷ | نقشه‌برداری جرائم به وسیله‌ی تکنولوژی GIS         |
| ۳۴۸ | گامی فراتر با مکان - تحلیل تخصیص                  |
| ۳۴۹ | استفاده از آمار فضایی پیچیده برای درک بهتر جرائم  |
| ۳۴۹ | ریاضیات فضایی پیشرفته                             |
| ۳۵۰ | آمار توصیفی                                       |
| ۳۵۲ | کاوش مشکلات با استفاده از علم داده در تحلیل جرائم |
| ۳۵۳ | توجه به موبایل‌های فنی                            |
| ۳۵۵ | بخش شش                                            |
| ۳۵۵ | آشنایی با ابزار ۰                                 |
| ۳۵۷ | فصل بیست و یکم                                    |
| ۳۵۷ | ده منبع پدیده برای داده‌های بار                   |
| ۳۵۹ | معرفی برنامه Data.gov                             |
| ۳۶۰ | بررسی داده‌های باز کانادا                         |
| ۳۶۱ | آشنایی با data.gov.uk                             |
| ۳۶۲ | بررسی داده‌های دفتر سرشماری ایالات متحده          |
| ۳۶۳ | معرفی داده‌های NASA                               |
| ۳۶۳ | گردآوری داده‌های بانک جهانی                       |
| ۳۶۵ | آشنایی با داده Knoema                             |
| ۳۶۶ | صف‌بندی با استفاده از داده‌های Quandle            |
| ۳۶۸ | بررسی داده‌های Exversion                          |
| ۳۶۹ | نگاشت داده‌های فضایی OpenStreetMap                |
| ۳۷۰ | منابع                                             |

# خط‌مشی کیفیت انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب‌هایی است که بتواند خواست‌هایی به روز جامعه فرهنگی و علمی کشور را تا حد امکان پوشش دهد.

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، مشارک واقع شویم.

گسترده‌گی علوه 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 اساسی‌ترین راه دستیابی به اطلاعات و اطلاع‌رسانی، بیش از پیش روشن می‌نماید. در این راستا، واحد انتشارات موسسه فرهنگی هنری دیباگران تهران با همکاری جمعی از اساتید، مؤلفان، مترجمان، متخصصان، پژوهشگران، محققان و نیز پرسنل ورزیده و ماهر در زمینه امور نشر درصدد هستند تا با تلاش‌های مستمر خود برای پاسخگویی به نیازهای موجود، منابعی پربار، معتبر و با کیفیت مناسب در اختیار علاقمندان قرار دهند.

کتابی که در دست دارید با همت "آقایان محمد جراد جعفری - پرمان محمد علیزاده" و تلاش جمعی از همکاران انتشارات میسر گشته که شاید به استایکایک این گرامیان تشکر و قدردانی کنیم.

## کارشناسی و نظارت بر محتوا: زهره قزلباش

در خاتمه ضمن سپاسگزاری از شما دانش‌پژوه گرامی درخواست می‌نمایم با مراجعه به آدرس [dibagaran.mft.info](mailto: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  
[Publishing@mftmail.com](mailto:Publishing@mftmail.com)



## مقدمه مولف

مرور زندگی دانشمندان بزرگ چه در قرون اخیر و چه دورتر منبع الهام بخش مناسبی برای انسان امروز است. نکته ای که به وفور در این بررسی ها مشاهده می شود منابع اطلاعاتی و نوع دسترسی به آن ها توسط انسان است. تعدادی کتاب از گذشتگان یک علم خاص، مقداری مشاهدات محدود و در ادامه تحلیل و نتیجه گیری و ثبت نتایج برای استفاده آیندگان. چالش در این موضوع داشتن منابع اطلاعاتی بیشتر چه از طریق مشاهده و چه از طریق متن های نوشته شده بود. شاید شما هم داستان فردی را شنید باشید که برای تهیه کتاب گرسنگی می کشید.

برگردیم به عصر حاضر. سمت تنظیمات حافظه کامپیوتر رومیزی، تلفن همراه هوشمند و یا سایر تجهیزات همراه خود ناهمی اندازهید. احتمالاً تعداد زیادی کتاب با موضوعات مختلف در دسترس شما هست. اگر اهل مطالعه باشید تعداد ممکن است به صدها عنوان هم برسد. می توانید سری به کتابخانه های اطراف خود بزنید. کتاب سبج ساده در گوگل در رابطه با عنوانی خاص داشته باشید.

دایره المعارف های بسیار بزرگ و کتابخانه های آنلاین را فراموش نکنید! شاید به دلایلی حوصله مطالعه کتاب های قطور را نداشته باشید. بیایید از فضاء کتاب و کتاب خواندن خارج شویم. امروز ها و کینز جدیدترین نظریه خود را در صفحه شخصی اش در فضای مجازی منتشر کرد. نظریکی از هم کلاسی های دوران دانشگاه را در مورد قسمت جدید فیلم جنگ ستارگان در صفحه مربوط به این فیلم خواندم.

خوب بهتر است کمی به کارهای شخصی خود برسیم. با دوستی که هر روزه با او از سال ها به دیدنش بروم توسط سیستم آدرس دهی گوگل نقشه محل قرار را مشخص کردیم. هر چند سیستم پیشنهاد دهنده گوگل به ما یک رستوران خوب معرفی کرد. چند عکس یادگاری با گوش هوشمند خود انداختیم و با سایر دوستان خودمان از طریق اینستاگرام به اشتراک گذاشتیم. البته آن روز همه چیز به خوبی و خوشی پیش نرفت و مجبور شدیم در نهایت به علت مسمومیت غذایی ناشی از حساسیت دوستم به بیمارستان با استفاده از نقشه گوگل برویم.

می توان ساعت ها از حجم بسیار عظیم منابع اطلاعاتی که ثانیه به ثانیه در حال ذخیره شدن و استفاده است صحبت کرد. اطلاعات زیادی از ما در حال ذخیره شدن و اطلاعات زیادی از بیرون در

معروض استفاده ما قرار گرفته اند. همه ی مواردی که گفته شد امروزه نقش کتاب ها و مشاهدات کمیاب گذشته را بازی می کنند که انسان باید برای بدست آوردنشان گرسنگی می کشید.

این مقدار از منابع اطلاعاتی و نیاز به مدیریت درست آن ها چالش جدیدی به نام استفاده از اطلاعات را بوجود آورد. در پس این چالش ها "علم داده" یا به عرصه زندگی گذاشت. سال ها پیش آقایان توماس دونپورت و دی جی پاتیل در سال ۲۰۱۲ در مقاله «علم داده: جذاب ترین شغل قرن بیست و یکم» متخصصین علم داده را این طور تعریف می کنند: کسانی که می دانند چگونه می توان از انبوه اطلاعات بدون ساختار پاسخ سوالهای کسب و کار را پیدا کرد. استنتون در سال ۲۰۱۳ علم داده را این طور تعریف می کند: علم داده رشته در حال ظهوری است که به جمع آوری، آماده سازی، تحلیل، بصره سازی، مدیریت و نگهداشت اطلاعات در حجم بالا می پردازد. در اسکول در سال ۲۰۱۴ علم داده را این سور تعریف می کند: علم داده مهندسی عمران داده هاست.

هدف کلی این کتاب آشنا ازی خواننده با دنیای "علم داده" و موضوعات مرتبط با آن است. خواننده پس از مطالعه این کتاب شناخت کلی با زوایای مختلف این دنیای بزرگ خواهد داشت تا پس از مطالعه هر بخش بتواند مسیر خود را در راه تبدیل شدن به یک متخصص داده به درستی دنبال کند. برای افراد غیر مبتدی نیز بخش های کتاب بصورتی است که می توانند بخش های مرتبط با حوزه کاری خود را مطالعه کنند هر چند خ نند همه بخش ها جهت داشتن دید بهتر از علم داده مطمئنا خالی از لطف نخواهد بود. در این کتاب پس از بررسی زوایای مختلف این علم به سه بخش اصلی آن می پردازیم که شامل آماده سازی، تحلیل و تصویر سازی اطلاعات می باشد. علم آمار مطمئنا بخشی مهم و جدایی ناپذیر در علوم مرتبط با داده ها است که در این کتاب توجه ویژه ای به آن شده است. لابلادر کنار مطالب مطرح شده در هر ضمیمه ابزارها و زبان های برنامه نویسی کاربردی همراه با مختصری معرفی از نحوه استفاده آن ها عنوان شده است. در فصل های پایانی نگاهی به کاربرد علم داده در محیط زیست، کشف جرائم، خبرنگاری و تجارت الکترونیک انداخته ایم.

این کتاب مقدمه ایست جهت شروع یک راه طولانی که درباره جاده ها، ابزارهای مورد نیاز، مقصد و راه مواجهه با هر پستی و بلندی را به شما نشان می دهد که برای لذت بردن و داشتن انرژی و انگیزه کافی در هر سفر دانستن این موارد لازم است. پس از این کتاب در آینده ای نزدیک بصورت خاص سفر خود را با کتاب "کاربرد پایتون در علم داده" ادامه خواهیم داد.

بدیهیست این کتاب خالی از اشکال چه در محتوی و چه در بخش های فرمی نیست و از اسانید و خوانندگان محترم درخواست می کنیم نظرات و اشکالات خود را از طریق آدرس های زیر با ما در میان بگذارند.

[https://t.me/Data\\_Science\\_i](https://t.me/Data_Science_i)

<https://www.linkedin.com/in/prmma۲۳>

این عربی در کتاب فصوص الحکم جهان را زانیده ی تصور زیبای خداوند می داند. بی شک اگر امروز او زنده بود علم داده را وسیله ی ارزیابی و کشف حقایق و زیبایی های این تصور الهی می دانست.

یا حق

مهندس محمد ج. د جعفری کارشناس ارشد فناوری اطلاعات و مدرس دانشگاه

www.ketab.ir