مؤسسه فرهنگی هنری
دیباگران تهران

به نام خدا

ملزومات آماری در علم داده ها

شامل بیش از ۵۰ کد R و پایتون

مؤلفان:

ساره هرمزان (کارشناسی ارشد علوم کامپیوتر از دانشگاه تهران)

محمد رضاپور (عضو هیأت علمی وزارت علوم ، تحقیقات و فناوری)

سینا توکلی بنیزی (کارشناسی مهندسی صنایع)

مؤسسه فرهنگی هنری
دیباگران تهران

عنوان کتاب: ملزومات آماری در علم داده ها
شامل بیش از ۵۰ کد R و پایتون

مولفان : ساره هرمزان- محمد رضاپور-سینا توکلی بنیزی

ناشر: موسسه فرهنگی هنری دیباگران تهران
ویراستار: نرگس مهربد
صفحه آرایی:نازنین نصیری
طراح جلد:داریوش فرسایی
نوبت چاپ: اول
تاریخ نشر: ۱۴۰۱
چاپ و صحافی:صدف
تیراژ:۱۰۰ جلد
قیمت:۲۴۰۰۰۰۰ ریال
شابک: ۵-۶۰۰-۲۱۸-۶۲۲-۹۷۸

سرشناسه:هرمزان،ساره،۱۳۶۸-
عنوان و نام پدیدآور: ملزومات آماری در علم داده ها: شامل بیش از ۵۰ کد R و پایتون/مولفان: ساره هرمزان، محمد رضاپور، سینا توکلی بنیزی؛ ویراستار: نرگس مهربد.
مشخصات نشر: تهران : دیباگران تهران :۱۴۰۱
مشخصات ظاهری:۳۵۴ ص:مصور،جدول،نمودار
شابک: ۵-۶۰۰-۲۱۸-۶۲۲-۹۷۸
وضعیت فهرست نویسی: فیپا
یادداشت: ص.ع. به انگلیسی Sareh Hormozan, Mohammad Rezapour, Sina Tavakoli
Essentials of statistical in data science with more than ۵۰ codes in R and Python.
یادداشت: کتابنامه: ص. ۳۵۲-۳۵۳
موضوع: آمار ریاضی- داده پردازی
موضوع: mathematical statistics- data processing
موضوع:نرم افزار آر (R(computer software
موضوع: پایتون(زبان برنامه نویسی کامپیوتر)
موضوع: python(computer program language)
شناسه افزوده:رضاپور،محمد،۱۳۵۷-
شناسه افزوده: توکلی بنیزی،سینا،۱۳۷۷-
رده بندی کنگره: QA ۲۷۶/۶
رده بندی دیویی: ۵۱۹/۵۰۲۸۵
شماره کتابشناسی ملی:۸۹۶۰۵۵۱

نشانی واحد فروش: تهران، میدان انقلاب، خ کارگر جنوبی، روبروی پاساژ مهستان، پلاک ۱۲۵۱-تلفن: ۶۶۴۱۰۰۴۶-۲۲۰۸۵۱۱۱
فروشگاههای اینترنتی دیباگران تهران :
WWW.MFTBOOK.IR
www.dibagarantehran.com

نشانی تلگرام:@mftbook نشانی اینستاگرام دیبا dibagaran_publishing

فهرست مطالب

فصل دوم

فصل سوم

فصل چهارم

فصل پنجم

فصل ششم

فصل هفتم

مقدمه ناشر

خط مشی انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب هایی با کیفیت عالی است که بتواند خواسته های 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.

گستردگی علوم وسرعت 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 زمینه های گوناگون و مورد نیاز جامعه تلاش نموده برای رفع کمبودها و نیازهای موجود، منابعی پُربار، معتبر و با کیفیت مناسب در اختیار علاقمندان قرار دهد.

کتابی که در دست دارید تألیف "**سرکارخانم ساره هرمزان و آقایان محمدرضاپور– و سینا توکلی بنیزی**" است که با تلاش همکاران ما در نشر دیباگران تهران منتشرگشته و شایسته است از یکایک این گرامیان تشکر و قدردانی کنیم.

با نظرات خود مشوق و راهنمای ما باشید

با ارائه نظرات و پیشنهادات وخواسته های خود،به ما کمک کنید تا بهتر و دقیق تر در جهت رفع نیازهای علمی و آموزشی کشورمان قدم برداریم.برای رساندن پیام هایتان به ما از رسانه های دیباگران تهران شامل سایتهای فروشگاهی و صفحه اینستاگرام و شماره های تماس که در صفحه شناسنامه کتاب آمده استفاده نمایید.

مدیر انتشارات

مؤسسه فرهنگی هنری دیباگران تهران
dibagaran@mftplus.com

پیش‌گفتاری بر «ملزومات آماری در علم داده‌ها»

روش‌های آماری بخش زیربناییِ علوم داده هستند؛ اما با این وجود، تعداد کمی از دانشمندان علوم داده آموزش رسمی در این زمینه دیده‌اند. از طرف دیگر، سرفصل‌های درس‌هایی؛ مانند آمار مهندسی یا آمار و احتمالات و کتاب‌های آماری نیز، به ندرت موضوع‌های آماری را برای تکمیل نیازمندی‌های یک دانشمند علوم داده‌ها و از نگرش داده‌کاوی پوشش می‌دهند. این کتاب، با پر کردن این جای خالی، فاصله ظاهراً قهرآمیز آمار و تحلیل داده‌ها با زبان‌های روز را پُر کرده و خواننده با شروع خواندن اولین فصل از کتاب درمی‌یابد، که تحلیلگران آماری فقط خود را به SPSS محدود نمی‌کنند و داده‌کاوان هم با وجود تسلط بر کدنویسی در زبان‌های تحلیل داده‌ها، از دانش آماری مربوطه بی‌نیاز نیستند.

این پیوند بین آمار و علوم داده‌ها را با مثال‌های جامعی در دو زبان R و Python به کتاب آورده‌ایم تا راهنمای کاربردی برای استفاده از روش‌های آماری برای علوم داده بوده و به شما کمک کند تا از این روش‌ها در جای مناسب استفاده کنید و بالأخره به شما نشان دهد که کدام روش مهم است و کدام‌یک از این منظر چندان اهمیتی ندارد!

بی‌آنکه بخواهیم زحمات نویسندگان دیگر را کم‌ارزش جلوه دهیم؛ اما این مهم را قاطعانه مطرح کرده و در پی رفع آن برآمدیم؛ اینکه بسیاری از منابع علوم داده از روش‌های آماری استفاده می‌کنند؛ اما فاقد چشم‌انداز آماری عمیق و درخور داوری علمی هستند. پس اگر شما تا کنون با یکی از زبان‌های R یا پایتون هم آشنایی دارید و از بعضی از آماره‌های آن استفاده کرده‌اید، این راهنمای سریع می‌تواند شکاف موجود را به‌طور خوانا و دسترس‌پذیر پُر نماید.

در این کتاب می‌آموزید:

- چرا تحلیل داده‌های توصیفی یک گام کلیدی در علوم داده است.
- چگونه نمونه‌گیری تصادفی می‌تواند بایاس را کاهش داده و به مجموعه داده‌ای باکیفیت‌تر حتی هنگام کار با کلان‌داده‌ها شود.
- چگونه اصول طراحی آزمایشی به جواب‌های قطعی برای سوالات منجر می‌شود.
- چگونه از رگرسیون برای تخمین خروجی‌ها و تشخیص ناهنجاری‌ها استفاده می‌شود.
- روش‌های یادگیری ماشین آماری که از روی داده‌ها «یاد می‌گیرند».
- روش‌های یادگیری نظارت‌نشده برای استخراج مفاهیم از داده‌های بدون برچسب.

این کتاب نه یک کتاب آمار و نه یک راهنمای یادگیری ماشین است، بلکه تلفیقی از این دو است: مفاهیم مهم و ضروری آماری با اصطلاحات داده‌کاوی که در زبان‌های امروزیِ تحلیل داده‌ها متداول هستند، در این کتاب تلفیق و تبیین می‌شوند، آن‌هم همراه با توضیحات واضح و مثال‌های فراوان. بی‌اغراق بگوییم، متنی آماده شده که برای مبتدیان علوم داده لازم و برای افراد باتجربه هم لذیذ است!

لذت کتاب صرفاً به خاطر مرور دانسته‌های دانشی نیست؛ کتاب مبنای آماری دقیق از کاربردهای مهندسی داده‌ها را شرح می‌دهد که یکی از فواید آن بسنده نکردن به خروجی نرم‌افزارها و ارائه پژوهش‌های قابل پذیرش در داوری‌های سخت‌گیرانه است!

در واقع، یکی از مشکلات سال‌های اخیر که گریبان‌گیر پژوهشگران شده، کار با نرم‌افزارهای محبوبی است که تحلیلگری داده را انجام داده و افراد بیش از آنکه به زیربنای آماری الگوریتم‌ها توجه داشته باشند، به نتیجه فوری آن ها بسنده می‌کنند و نتیجه این فرایند، عدم پذیرش مقالات برخی از فارغ‌التحصیلان عزیز ما که روی پای خود ایستاده‌اند، توسط داوران گرامی و متبحر برخی مجلات معتبر گردیده‌است و این کتاب بخش قابل توجهی از این معضل را با ترویج «علم مفید» جبران می‌نماید و امید است گام کوچکی در جهت تعبیرِ این دعای پُرمعنای پیامبر گرامی(ص) باشد که:

«اللهم إنّی أعوذ بکَ مِن عِلمٍ لایَنفَع»

«خدایا به تو پناه می‌برم از علمی که سودی نمی‌بخشد!».

مخاطبان کتاب را می‌توان دانشجویان مقاطع کارشناسی و ارشد رشته‌های علوم کامپیوتر، فناوری اطلاعات، مهندسی نرم‌افزار، مهندسی صنایع، مدیریت فناوری، ریاضی کاربردی، آمار، مدیریت و تمامیِ علاقمندان به مباحث آماری و تحلیلگری داده‌ها دانست، که ضمن تقدیم مطالب به خوانندگان گرامی، درخواست دارد ما را از نظرات سازنده خویش دریغ نفرمایند.