



به نام خدا

آموزش تحلیل داده با

JULIA

مؤلف:

احمد چیت ساززاده



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 می‌گیرند.

◀ عنوان کتاب: آموزش تحلیل داده با JULIA

◀ مولف: احمد چیت ساززاده

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ ویراستار: نرگس مهربد

◀ صفحه آرای: شبنم هاشم زاده

◀ طراح جلد: داریوش فرسای

◀ نوبت چاپ: اول

◀ تاریخ نشر: ۱۴۰۱

◀ چاپ و صحافی: صدف

◀ تیراژ: ۱۰۰ جلد

◀ قیمت: ۱۸۲۰۰۰۰ ریال

◀ شابک: ۹۷۸-۶۲۲-۲۱۸-۵۹۳-۰

نشانی واحد فروش: تهران، میدان انقلاب،

خ کارگر جنوبی، روبروی پاساژ مهستان،

پلاک ۱۲۵۱-تلفن: ۰۴۶-۶۶۴۱۰۰۶-۲۲۰۸۵۱۱۱

فروشگاههای اینترنتی دیباگران تهران :

WWW.MFTBOOK.IR

www.dibagaran-tehran.com

سرشناسه: چیت ساز زاده، احمد، ۱۳۷۳-

عنوان و نام پدیدآور: آموزش تحلیل داده با JULIA

/مولف: احمد چیت ساز زاده.

مشخصات نشر: تهران: دیباگران تهران: ۱۴۰۱

مشخصات ظاهری: ۲۵۶ ص: مصور.

شابک: ۰-۵۹۳-۲۱۸-۶۲۲-۹۷۸

وضعیت فهرست نویسی: فیپا کتابنامه.

موضوع: جولیا (زبان برنامه نویسی کامپیوتر)

موضوع: Julia (computer program language)

موضوع: زبان های نوشتاری (کامپیوتر)

موضوع: Scripting language (computer science)

موضوع: ساختار داده ها

موضوع: Data structures (computer science)

رده بندی کنگره: ۷۶/۷۳ QA

رده بندی دیویی: ۰۰۵/۷۳

شماره کتابشناسی ملی: ۸۹۶۵۱۲۲

نشانی تلگرام: @mftbook نشانی اینستاگرام دیبا dibagaran_publishing

هر کتاب دیباگران، یک فرصت جدید علمی و شغلی.

هر گوشی همراه، یک فروشگاه کتاب دیباگران تهران.

از طریق سایتهای دیباگران، در هر جای ایران به کتابهای ما دسترسی دارید.

فهرست مطالب

فصل اول

مقدمه

- 11 _____ پیش از هر چیز
- 11,1 _____ تحلیل داده؟ جولیا جواب می‌دهد!
- 12 _____ 1,2 کمی در مورد داده‌ها
- 15 _____ 1,3 کدزنی را که همه بلدند، برنامه‌نویس باش!

فصل دوم

تحلیل؛ یک بازی قدیمی در عصر جدید

- 18 _____ 2,1 قوانین بازی
- 19 _____ 2,2 مراحل بازی
- 21 _____ 2,3 انتخاب سلاح

فصل سوم

سلاح جولیا؟ چرا؟!

- 23 _____ 3,1 سخنی با تازه‌کاران
- 24 _____ 3,2 برای برنامه‌نویسان

26 3,3 قاموساً جولیا می‌تونه؟! _____

36 3,4 جولیا انتخاب غول‌ها! _____

فصل چهارم

قلق‌گیری جولیا

39 4,1 محیط‌های توسعه _____

42 4,2 نحو یا روال _____

59 4,3 ساختارهای داده جولیاپی _____

101 4,4 سامانه مدیریت فایل _____

102 4,5 کتابخانه‌های استاندارد هسته جولیا _____

فصل پنجم

کار با دیتافریم‌ها توسط کتابخانه DataFrames.jl

128 5,1 ذخیره‌سازی و فراخوانی فایل‌ها در جولیا _____

136 5,2 جایگاه‌یابی و خلاصه‌سازی داده‌ها _____

140 5,3 فیلتر کردن و زیرمجموعه‌گیری _____

147 5,4 نحو select() برای انتخاب ستونی _____

153 5,5 داده‌ها و تایپ‌های گمشده _____

159 5,6 پیوند دادن _____

169 5,7 تبدیلات متغیرهای دیتافریم _____

175 5,8 گروه‌بندی و ترکیب داده‌ها _____

180 _____ 5,9 بازدهی بهینه

فصل ششم

تصویرسازی داده‌ها با Makie.jl

191 _____ CairoMakie.jl مکانیزم 6,1

192 _____ دستکاری ویژگی‌های نمودار 6,2

201 _____ کار با تم‌ها 6,3

208 _____ استفاده از LaTeXStrings.jl 6,4

211 _____ رنگ‌ها و نقشه‌های رنگ 6,5

220 _____ تغییر در چینش‌ها 6,6

238 _____ GLMakie.jl مکانیزم 6,7

253 _____ منابع

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 دست‌دارید تألیف "جناب آقای احمد چیت ساززاده" است که با تلاش همکاران ما در نشر دیباگران تهران منتشر گشته و 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 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

خوب به خاطر دارم سال 96 که با جولیا آشنا شدم کاملاً شگفت‌زده و در عین حال شکاک بودم! شگفت‌زده به خاطر مزایای رقابتی جولیا و شکاک به خاطر اینکه آیا واقعاً آنچه که ادعا می‌کنند، وجود دارد؟ از آنجایی که ناآشنا با زبان‌های دیگر نبودم و پایتون نیز زبان اصلی‌ام بود، شروع به مطالعه و آزمایش کردم و در کمال ناباوری ادعاهای جولیا یکی پس از دیگری برایم ثابت شد که برخی از آن‌ها در فصل سوم آمده‌اند.

اما چرا همیشه برای یادگیری تازه‌واردان به حوزه برنامه‌نویسی، پایتون پیشنهاد می‌شود؟ پاسخ بسیار ساده است: «چون پایتون در عین کاربردهای بسیارش، از همه ساده‌تر است و مهم‌تر از همه منابع پایتون به زبان پارسی بیش‌ترند» و این شاید دلیلی باشد بر این پرسش تازه‌واردن مبنی بر اینکه «بهترین زبان برنامه‌نویسی کدام است؟» چرا که منابع فقط نحو¹ را آموزش می‌دهند بدون آنکه در ابتدا بگویند؛ «اگر می‌خواهید از سوپرمارکت سر کوچه خرید کنید، با دوچرخه یا نهایتاً با موتور بروید و اصلاً نیازی نیست که پورشه یا کامیونت خود را برای این کار استفاده کنید».

تمامی این‌ها دست به دست هم می‌دهند تا نیروی انسانی تازه‌نفس و با انگیزه را یا تلف کنند یا پس از آنکه خبره شدند، خبر از آن دهند که مسیر را اشتباه رفتی! و دنیا به سمت دیگری در حال حرکت است؛ برای همین اکثر برنامه‌نویسان، پایتون را نیز در انبار دانش خود دارند، هر چند که هیچ‌گاه نیازی به آن نداشته‌اند.

راه‌حل چیست؟ برای پیدا کردن راه‌حل باید ابتدا رفتارشناسی انجام داد و یک نتیجه کلی از رفتارشناسی تازه‌واردان این است، که آن‌ها نمی‌دانند برای چه برنامه‌نویسی کردن را نیاز دارند و زیاد میانه خوبی با جستجو به زبانی غیر از زبان مادری‌شان ندارند، که البته در ابتدای کار حق با آن‌هاست؛ بنابراین، راه‌حل این است که برنامه‌نویسان باتجربه‌تر به تکاپو افتاده و هر کدام، ابزارهای حوزه خودشان

¹ Syntax

را در جامعه رواج دهند. با این کار هم انبار دانش خودشان پُربارتر می‌شود هم تازه‌نفسان، سردرگم نمانده و اشتباه انتخاب نمی‌کنند.

حال تصور کنید در آن زمان که شروع کردم، مطالب زیادی به زبان پارسی درمورد جولیا وجود نداشت. خب این به ضرر منابع انسانی علمی یک کشور تمام می‌شد (باور کنید اغراق نمی‌کنم؛ چرا که طی این چندسال، اکثر کسانی که به بنده برای یادگیری جولیا مراجعه می‌کنند، دانشجویان ارشد و دکتری علوم پایه، انسانی و مهندسی بوده‌اند). می‌پرسید چرا؟! چون دنیای علمی، امروزه به سرعتی فراتر از پایتون نیاز دارد و همزمان سادگی نحو (سینتکس) آن را نیز می‌خواهد و این دقیقاً آن چیزی است که جولیا فراهم کرده و در فصل سوم به تفصیل درمورد آن سخن خواهیم گفت. خلاصه سخن آنکه بنابر تمامی دلایل بالا تصمیم گرفتیم معرفی و آموزش دادن زبان برنامه‌نویسی جولیا را به صورتی جدی پیگیری کنیم و این کتاب، ادامه راه من برای تحقق همان تصمیم است.

اما چرا جولیا با چاشنی تحلیل داده را انتخاب کردم؟ خب بسیار واضح است! همانطور که گفتم برنامه‌نویسان باتجربه‌تر باید ابزارهای مورد استفاده در حوزه خودشان را معرفی کنند تا علاوه بر آشنایی تازه‌واردان به این حوزه، تجربه‌داران نیز با چالش‌های مهم‌تری سروکله بزنند، به جای آنکه وقت خود را با چالش‌های پیش‌پافتاده تلف کنند.

علاوه بر این‌ها جولیا زبانی است که برای علم و محاسبات آن آفریده شده‌است؛ بنابراین، چه بهتر که در مسیر یادگیری این زبان شگفت‌انگیز به سمتی حرکت کنیم که در پایان مسیر علاوه بر فراگرفتن جولیا، یکی از کاربردهای عملی‌اش را که برای آن ساخته شده‌است نیز در انبار دانش خود داشته باشیم. حال چه کاربردی بهتر از تحلیل داده، در حالی که تصمیمات دنیا بر اساس داده‌های موجود و تحلیل آن‌ها به دست می‌آید؟! البته اگر هنوز شک دارید که هر شخصی باید در دنیای آینده که چندان هم دور نیست، تحلیل داده را در انبار دانشش داشته باشد، فصل اول و دوم در پاسخ به شماست.

کتاب پیش رو حاصل تحقیقات و ارائه‌های جامعه حامی نرم‌افزارهای منبع آزاد؛ خصوصاً دوستداران زبان برنامه‌نویسی جولیا می‌باشد. تلاش‌هایی که سعی کرده‌اند انحصارطلبی را شکسته و عدالت دسترسی و دانش را به علاقه‌مندان و فعالان دنیای دیجیتال هدیه کنند؛ بنابراین، با کتابی روبه‌رو هستید که صرفاً برای فروش نوشته نشده و به اصطلاح زرد نیست. این کتاب بر مبنای مقالات و کتاب‌های علمی و مستندات زبان‌های برنامه‌نویسی؛ چون پایتون، R و به ویژه جولیاست.

در این کتاب سعی شده حداقل اشتباهات علمی و انتقالات مفهومی به کار گرفته شود؛ اما هیچکس نمی‌تواند ادعا کند که بدون نقص است و بنده نیز از این قاعده مستثنی نیستم. بنابراین، بسیار خوشحال می‌شوم تا منت بر سر بنده گذاشته و اشتباهات یا انتقال بد مفاهیم را گوش‌زد کنید تا علاوه بر لطفی که به بنده خواهید نمود، در ویرایشات بعدی آن‌ها را اصلاح کرده و این نوشتار را به منبعی بدون نقص برای جامعه علمی نه تنها کشور عزیزمان ایران، بلکه تمام پارسی‌زبانان جهان تبدیل کنیم. در پایان بر خود واجب می‌دانم از جناب دکتر رضا ملکی¹ و جناب دکتر علی سوفالی² که باعث شدند بنده در این مسیر قرار گرفته و با راهنمایی‌های فوق‌العاده‌شان شیرینی علم داده را بچشم، کمال تقدیر و تشکر را داشته باشم.

همچنین متشکرم از نویسندگان کتاب Julia Data Science که بی‌منت اجازه اقتباس و کپی‌برداری از کتابشان را نه تنها به بنده که به تمامی جامعه طرفداران لایسنس‌های آزادسازی داده‌اند و کتاب حال حاضر اقتباسی از آن است.

محل شروع بازی: برای دانستن چیرستی این محل باید کتاب را به انتها برسانید.

امید به علم‌افروزی هر چه بیش‌تر
تقدیم به پدر، مادر و همسر عزیزم
احمد چیت‌ساززاده
28 تیرماه 1401

¹<https://www.linkedin.com/in/reza-maleki-36106693>

²<https://www.linkedin.com/in/ali-sufali>