

راهنمای کاربردی پایتون برای علم داده‌ها

ب

۱

۱

۱

۱

نویسندگان

علیرضا افضل آقایی نائینی

کوروش پرند

www.ketab.ir



سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران

سرشناسه	: افضل آقای نائینی، علیرضا، ۱۳۷۵
عنوان و نام پدیدآور	: راهنمای کاربردی پایتون برای علم داده‌ها / نویسندگان علیرضا افضل
مشخصات نشر	: آقای نائینی؛ کورش پرند تهران، آرون، ۱۴۰۲
مشخصات ظاهری	: ۳۸۰ ص.
شابک	: ۹۷۸ - ۶۲۲ - ۳۱۷ - ۰۹۶ - ۶
وضعیت فهرست‌نویسی	: فیا
موضوع	: پایتون (زبان برنامه نویسی کامپیوتر) Python (Computer program language)
شناسه افزوده	: پرند، کورش، ۱۳۵۰
شناسه افزوده	: جلالی، مریم، ۱۳۵۲، ویراستار
رده‌بندی کنگره	: QA ۷۶ / ۷۳
رده‌بندی دیویی	: ۰۰۵ / ۱۳۳
شماره کتابخانه ملی	: ۹۲۰۷۵۳۳



راهنمای کاربردی پایتون برای علم داده‌ها

نویسندگان: علیرضا افضل آقای نائینی، کورش پرند

طراح جلد: محبوبه افضل آقای نائینی

ناشر: انتشارات آرون

چاپ اول: ۱۴۰۲

چاپ صدف: ۵۰ نسخه

۳۵۰/۰۰۰ تومان

نشانی: میدان انقلاب، خیابان ۱۲ فروردین، خیابان وحید نظری

نرسیده به خیابان منیری جاوید، پلاک ۱۰۵، واحد ۳ تلفن: ۵۱ - ۶۶۹۶۲۸۵۰

ایمیل: Arvannashr@yahoo.com وبسایت: www.Arvanashr.ir

۹۷۸ - ۶۲۲ - ۳۱۷ - ۰۹۶ - ۶

فهرست مطالب

فصل ۱ آشنایی با پایتون

۱۳

- ۱-۱ چرا پایتون؟ ۱۳
- ۱-۲ پایتون چیست؟ ۱۴
- ۱-۳ نحوه نصب ۱۵
 - ۱-۳-۱ ویندوز ۱۵
 - ۱-۳-۲ گنو/لینوکس ۱۷
 - ۱-۳-۳ مکینتاش ۱۸
 - ۱-۳-۴ کامپایل و نصب از کد منبع ۱۹
- ۱-۴ نرم افزارها و ابزارهای کمکی ۱۹
 - ۱-۴-۱ آناکوندا ۱۹
 - ۱-۴-۲ PyEnv ۱۹
- ۱-۵ محیط های برنامه نویسی ۲۰
 - ۱-۵-۱ نوت بوک ژوپیتر ۲۱
 - ۱-۵-۲ محیط های برنامه نویسی آنلاین ۲۴
- ۱-۶ منابع برخط ۲۶

فصل ۲ نحو و نگارش

۲۷

- ۲-۱ کلمات کلیدی ۳۰
- ۲-۲ متغیرها، انواع داده ۳۲
- ۲-۳ ساختارهای داده اولیه ۳۵
- ۲-۴ ساختارهای داده مرکب ۳۸
 - ۲-۴-۱ لیست ۳۹

۴۲	چندتایی	۲-۴-۲
۴۲	رشته	۲-۴-۳
۴۵	۱.۲-۴-۳ قالب‌دهی رشته	
۵۲	مجموعه	۲-۴-۴
۵۳	دیکشنری	۲-۴-۵
۵۵	اندیس‌گذاری و برش	۲-۴-۶
۶۳	گشودن چندتایی‌ها	۲-۴-۷
۶۵	گزارمان	۲-۵
۶۶	عملگرها و عملوندها	۲-۶
۶۹	۲-۶-۱ اعداد صحیح، اعشاری و مختلط	
۷۲	۲-۶-۲ لیست، چندتایی و رشته	
۷۴	۲-۶-۳ مجموعه	
۷۵	۲-۶-۴ دیکشنری	
۷۶	۲-۷ توابع	
۷۷	۲-۷-۱ توابع توکار	
۹۱	۲-۷-۲ تعریف توابع	
۱۰۰	۲-۷-۳ توابع مولد	
۱۰۱	۲-۷-۴ حوزه متغیرها	
۱۰۴	۲-۷-۵ آرایه‌گرها	
۱۰۵	۲-۸ شی‌گرایی	
۱۱۱	۲-۸-۱ وراثت	
۱۱۱	۲-۸-۲ کپسوله‌سازی	
۱۱۲	۲-۸-۳ انتزاع	
۱۱۳	۲-۸-۴ چندریختی	
۱۱۳	۲-۹ کتابخانه‌ها	
۱۱۶	۲-۱۰ ساختارکنترلی شرط	
۱۱۶	۲-۱۰-۱ دستور if	
۱۱۹	۲-۱۰-۲ تطابق	
۱۲۱	۲-۱۱ ساختارکنترلی تکرار	
۱۲۲	۲-۱۱-۱ حلقه while	
۱۲۵	۲-۱۱-۲ حلقه for	
۱۲۵	۱.۲-۱۱-۲ پیمایشگرها	

۱۳۱	۲-۱۲ بلوک‌های تودرتو
۱۳۶	۲-۱۲-۱ عبارات کنترل حلقه
۱۳۸	۲-۱۳ مدیریت خطا
۱۳۸	۲-۱۳-۱ استثناها
۱۴۰	۲-۱۳-۲ مدیریت استثناها
۱۴۲	۲-۱۴ برنامه‌نویسی ناهم‌زمان

فصل ۳ کتابخانه‌های کاربردی ۱۴۷

۱۴۸	۳-۱ محیط مجازی
۱۴۹	۳-۱-۱ venv
۱۵۰	۳-۱-۲ conda
۱۵۱	۳-۲ آمار و ریاضی
۱۵۱	۳-۲-۱ کتابخانه math
۱۵۲	۳-۲-۲ کتابخانه random
۱۵۵	۳-۲-۳ کتابخانه numpy
۱۶۴	۳-۲-۴ کتابخانه scipy
۱۶۶	۳-۲-۵ کتابخانه sympy
۱۶۷	۳-۲-۶ کتابخانه distfit
۱۶۸	۳-۳ زمان و تاریخ
۱۶۸	۳-۳-۱ کتابخانه time
۱۷۰	۳-۳-۲ کتابخانه datetime
۱۷۲	۳-۳-۳ کتابخانه jdatetime
۱۷۳	۳-۴ کار با داده و فایل
۱۷۳	۳-۴-۱ کتابخانه csv
۱۷۶	۳-۴-۲ کتابخانه json
۱۸۲	۳-۴-۳ کتابخانه beautifulsoup
۱۸۳	۳-۴-۴ کتابخانه pandas
۱۹۰	۱.۳-۴-۴ پاک‌سازی داده‌ها
۱۹۱	۲.۳-۴-۴ یکپارچه‌سازی داده‌ها
۱۹۱	۳.۳-۴-۴ کاهش داده‌ها
۱۹۲	۴.۳-۴-۴ تبدیل داده‌ها
۱۹۴	۳-۴-۵ کتابخانه glob

پیش‌گفتار

علم داده‌ها یک زمینه علمی بین رشته‌ای است که در آن ریاضیات، آمار، برنامه‌نویسی کامپیوتری، هوش مصنوعی، داده‌کاوی و یادگیری ماشین با یک تخصص خاص ترکیب می‌شوند تا اطلاعات و ساختارهای پنهان در مجموعه‌ای از داده‌ها استخراج شوند. از اطلاعاتی که به دست آمده است می‌توان در تصمیم‌گیری‌های مهم استفاده کرد. از آنجایی که امروزه با حجم بالایی از داده‌ها روبه‌رو هستیم، اهمیت علم داده‌ها پررنگ‌تر می‌شود. درواقع می‌توان گفت امروزه یکی از زمینه‌های مهم در دنیا که برای اغلب سازمان‌ها ضروری است علم داده‌ها است. از این‌رو بسیاری از شرکت‌های بزرگ متخصصانی در این زمینه دارند و یا مشغول آموزش کارمندان خود در این زمینه هستند.

در دوره‌های برگزار شده توسط مؤلفان این کتاب، مباحث اساسی در ارتباط با علم داده‌ها جهت آمادگی یک فرد برای ورود به بازار کار یا محیط‌های آموزشی، مطرح و تدریس شده‌اند. این مباحث پودمان‌های پیش‌پردازش داده‌ها، آموزش پایتون، رگرسیون و دسته‌بندی، خوشه‌بندی، الگوریتم‌های فرااکتشافی و یادگیری عمیق را در بر می‌گیرند.

در میان ابزارهای پیاده‌سازی الگوریتم‌های علم داده، زبان برنامه‌نویسی پایتون به یکی از قدرت‌های بی‌بدیل تبدیل گشته است. سادگی نحو این زبان به همراه وجود کتابخانه‌های توکار و شخص ثالث فراوان آن‌را به انتخاب اول شرکت‌ها، محققان و دانشمندان علم داده تبدیل کرده است. تجربه نویسندگان این کتاب از سال‌ها تدریس این زبان برنامه‌نویسی در محیط‌های آموزشی و شغلی نشان داده است که مخاطبین به سادگی با مفاهیم و نحو این زبان ارتباط برقرار می‌کنند و گاهی با شور و هیجان بالایی خواهان پیگیری مفاهیم گفته شده هستند.

هرچند که کتاب‌های مختلفی در سالیان گذشته در این حیطه به تحریر درآمده‌اند، اما در نگارش این کتاب تلاش شده است از دیدگاه پروژه-محوری به این شاخه بپردازیم؛ به‌طوری که خواننده بتواند با خواندن آن توانایی پیاده‌سازی یک خط‌لوله برای پروژه علم داده خود، از جمع‌آوری داده و پیش‌پردازش، تا آموزش مدل، تحلیل و ایجاد یک نرم‌افزار کاربردی جهت پیش‌بینی را به دست آورد.

از این رو این کتاب در چهار فصل تدوین شده است. در فصل اول خواننده با مبانی و مفاهیم زبان‌های برنامه‌نویسی، علی‌الخصوص زبان پایتون آشنا خواهد شد. در فصل دوم به‌صورت خاص، نحو نسخه ۳/۱۱ این زبان برنامه‌نویسی را بیان خواهیم کرد. در فصل سوم نیز به معرفی ۸۰ کتابخانه پرکاربرد این زبان در

علم داده‌ها خواهیم پرداخت. این بسته‌های نرم‌افزاری با بررسی بیش از ۲۰۰۰ پروژه کاربردی علم داده‌ها در مخازن متن‌باز وب‌سایت گیت‌هاب، انتخاب شده‌اند. هرچند که برای هر کدام از این ابزارها می‌توان کتاب‌های متعددی تألیف نمود، اما رویکرد این کتاب آشنایی خواننده با انواع ابزارهای موجود در این زبان برنامه‌نویسی برای علم داده‌ها است. بنابراین شایسته است در صورت عدم آشنایی با مفاهیم مذکور به دیگر کتاب‌های این مجموعه مراجعه شود. در فصل پایانی کتاب نیز به کمک ابزارهای معرفی شده، به انجام چند پروژه کاربردی مرتبط با علم داده‌ها خواهیم پرداخت.

کتابی که پیش رو دارید حاصل سال‌ها مطالعه و تجربه مؤلفان آن است. نویسندگان این کتاب، پروژه‌ها و اختراعات متعددی را به کمک این زبان برنامه‌نویسی به سرانجام رسانده‌اند. همچنین تجربیات ایشان در برگزاری دوره‌های مختلف در زمینه‌های یادگیری ماشین، یادگیری عمیق، داده‌کاوی و آموزش زبان برنامه‌نویسی پایتون در مراکز علمی، اداری و صنعتی مختلف در این کتاب خلاصه شده است. بنابراین این کتاب نه تنها برای دانشجویان رشته‌های مختلف می‌تواند مفید باشد، بلکه پژوهشگران، مدیران و کارمندان با عناوین مختلف نیز می‌توانند از آن بهره‌مند شوند.

در نگارش این کتاب سعی شده است مطالب به ساده‌ترین شکل و زبان بیان شوند. قطعاً هیچ اثری نمی‌تواند جای پیشنهاد و انتقاد نداشته باشد. از این رو باعث خرسندی نویسندگان این کتاب خواهد بود تا با نظرات و پیشنهادات ارزشمند خود از طریق پست الکترونیکی alirezaafzalaghaci@gmail.com باعث بهبود مطالب این کتاب شوید.

با سپاس

علیرضا افصل آقایی و کورش پرند بهار ۱۴۰۲