



به نام خدا



علم داده در پایتون

مفاهیم و کاربردها

مؤلفان:

محسن یزدی نژاد (کандیدای دکتری هوش مصنوعی)

محمدرضا پور (استادیار، عضو هیئت علمی وزارت علوم، تحقیقات و فناوری)

معصومه جوادی نیا (کارشناسی مهندسی نرم افزار)

فرهاد استوار نژاد (کارشناسی ارشد حسابداری)



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.

عنوان کتاب: علم داده در پایتون، مفاهیم و کاربردها

مؤلفان: محسن یزدی نژاد - محمد رضاپور - معصومه جوادی نیا -

فرهاد استوارنژاد

ناشر: موسسه فرهنگی هنری دیباگران تهران

ویراستار: نرگس مهرید

صفحه آرای: نازنین نصیری

نوبت چاپ: سوم

تاریخ نشر: ۱۴۰۲

چاپ و صحافی: صدف

تیراژ: ۱۰۰ جلد

قیمت: ۲۰۰۰۰۰۰ ریال

شابک: ۹۷۸-۶۲۲-۲۱۸-۵۲۹-۹

نشانی واحد فروش: تهران، خیابان انقلاب، خ دانشگاه -

تقاطع شهدای زاندارمری - ساختمان اداری دانشگاه -

طبقه دوم - واحد ۴ تلفن: ۲۲۰۸۵۱۱۱-۶۶۴۹۸۱۶۸

فروشگاههای اینترنتی دیباگران تهران :

WWW.MFTBOOK.IR

www.dibagaran-tehran.com

عنوان و نام پدیدآور: علم داده در پایتون، مفاهیم و کاربردها/

مؤلفان: محسن یزدی نژاد و... دیباگران؛

ویراستار: نرگس مهرید.

مشخصات نشر: تهران: دیباگران تهران: ۱۴۰۲

مشخصات ظاهری: ۲۰۰ ص. مصور. جدول

شابک: ۹۷۸-۶۲۲-۲۱۸-۵۲۹-۹

وضعیت فهرست نویسی: فیبا

یادداشت: مؤلفان: محسن یزدی نژاد، محمد رضاپور، معصومه جوادی نیا، فرهاد استوارنژاد.

موضوع: پایتون (زبان برنامه نویسی کامپیوتر)

موضوع: Python (computer program language)

موضوع: زبان شناسی کامپیوتری computational linguistics

موضوع: داده کاوی Data mining

شناسه افزوده: یزدی نژاد، محسن، ۱۳۶۲-

رده بندی کنگره: QA ۷۶/۷۳

رده بندی دیویی: ۰۰۵/۱۲۳

شماره کتابشناسی ملی: ۸۸۰۵۲۴۶

نشانی اینستاگرام دیبا dibagaran_publishing

نشانی تلگرام: @mftbook

هر کتاب دیباگران، یک فرصت جدید علمی و شغلی.

هرگوشی همراه، یک فروشگاه کتاب دیباگران تهران.

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

فهرست مطالب

مقدمه ناشر ۷

مقدمه نویسندگان ۸

فصل اول: مقدمه‌ای بر داده‌کاوی ۱۱

۱-۱- تعریف داده‌کاوی ۱۲

۱-۲- چرا داده‌کاوی؟ ۱۳

۱-۳- هدف از داده‌کاوی ۱۴

۱-۴- جایگاه داده‌کاوی، هوش مصنوعی، یادگیری ماشین ۱۴

۱-۵- فرآیند داده‌کاوی CRISP ۱۶

۱-۶- وظایف داده‌کاوی ۱۸

فصل دوم: پیش‌پردازش داده‌ها ۲۰

۲-۱- مفاهیم اولیه در داده‌کاوی ۲۱

۲-۲- کیفیت داده ۲۲

۲-۲-۱- نویزها و داده‌های دورافتاده ۲۲

۲-۲-۲- روش‌های پیدا کردن نقاط دورافتاده ۲۳

۲-۲-۳- مقادیر از دست‌رفته ۲۵

۲-۲-۴- داده تکراری ۲۷

۲-۲-۵- داده‌های جهت‌دار ۲۷

۲-۳- آماده‌سازی داده‌ها ۲۷

۲-۳-۱- یکپارچه‌سازی ۲۸

۲-۳-۲- تجمیع ۲۸

۲-۳-۳- نمونه‌گیری ۲۸

۲-۳-۳-۱- نمونه‌گیری تصادفی ساده ۳۰

۲-۳-۳-۲- نمونه‌گیری سیستماتیک ۳۰

۲-۳-۳-۳- نمونه‌گیری چندبخشی (طبقه‌ای) ۳۱

۲-۳-۳-۴- نمونه‌گیری خوشه‌ای ۳۲

۲-۳-۴- کاهش بعد ۳۳

۲-۳-۴-۱- انتخاب ویژگی ۳۴

۲-۳-۴-۲- استخراج ویژگی ۳۶

۲-۳-۴-۳- خلق ویژگی ۳۷

۲-۳-۴-۴- گسسته‌سازی ۳۸

۲-۳-۴-۵- انتقال ویژگی ۳۸