



به نام خدا

**داده کاوی با نرم افزار**

**ORANGE**

مؤلفان:

**محمد فاتحی**

**مسعود کاویانی**



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: داده کاوی با نرم افزار ORANGE

◀ مولفان: محمد فاتحی، مسعود کاویانی

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ ویراستار: مهدیه مخبری

◀ صفحه آرای: فرنوش عبدالهی

◀ طراح جلد: داریوش فرسایی

◀ نوبت چاپ: اول

◀ تاریخ نشر: ۱۴۰۲

◀ چاپ و صحافی: صدف

◀ تیراژ: ۱۰۰ جلد

◀ قیمت: ۱۶۲۰۰۰۰ ریال

◀ شابک: ۹۷۸-۶۲۲-۲۱۸-۶۵۳-۱

◀ نشانی واحد فروش: تهران، خیابان انقلاب، خیابان دانشگاه

◀ تقاطع شهدای ژاندارمری - پلاک ۱۵۸ ساختمان دانشگاه -

◀ طبقه دوم - واحد ۴ تلفن ها: ۶۶۹۶۵۷۴۹-۲۲۰۸۵۱۱۱

◀ فروشگاههای اینترنتی دیباگران تهران:

[WWW.MFTBOOK.IR](http://WWW.MFTBOOK.IR)

[www.dibagaran-tehran.com](http://www.dibagaran-tehran.com)

سرشناسه: کاویانی، مسعود، ۱۳۶۹ - Kaviani, masoud  
عنوان و نام پدیدآور: داده کاوی با نرم افزار Orange  
/مولفان: مسعود کاویانی، محمد فاتحی؛  
ویراستار: مهدیه مخبری.  
مشخصات نشر: تهران: دیباگران تهران: ۱۴۰۱  
مشخصات ظاهری: ۱۵۲ ص: مصور،  
شابک: ۱-۶۵۳-۲۱۸-۶۲۲-۹۷۸  
وضعیت فهرست نویسی: فیپا  
موضوع: نرم افزار اورنج (orange (computer file  
موضوع: هوش مصنوعی Artificial intelligence  
موضوع: داده کاوی - نرم افزار data mining - software  
شناسه افزوده: فاتحی، محمد، ۱۳۶۶ -  
رده بندی کنگره: QA ۷۶/۹  
رده بندی دیویی: ۰۰۶/۳۱۲  
شماره کتابشناسی ملی: ۹۱۲۹۳۳۴

نشانی اینستاگرام دیبا [dibagaran\\_publishing](https://www.instagram.com/dibagaran_publishing) نشانی تلگرام: @mftbook

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 ۱ / مبانی و مقدمات داده کاوی ..... ۷

۸	..... داده کاوی چیست؟
۹	..... تفاوت هوش مصنوعی، علم داده، داده کاوی و یادگیری ماشین
۱۱	..... زیر حوزه های داده کاوی
۱۲	..... انواع ساختارهای مسائل در داده کاوی
۱۴	..... رویکرد یادگیری نظارت شده در مقابل یادگیری غیرنظارت شده
۱۷	..... تفاوت نمونه (مشاهده) و ویژگی (متغیر)
۱۸	..... متدولوژی CRISP در داده کاوی
۲۱	..... جبر خطی
۲۶	..... ساختار فصل های آینده این کتاب

## فصل ۲ / معرفی و نصب نرم افزار ORANGE ..... ۲۸

۲۹	..... مقدمه
۲۹	..... نصب و راه اندازی
۳۴	..... آشنایی با محیط نرم افزار

## فصل ۳ / آماده سازی و بصری سازی داده ها ..... ۴۸

۴۹	..... مقدمه
۵۰	..... پیش پردازش داده ها
۵۰	..... حذف مقادیر گم شده
۵۴	..... نرمال سازی داده ها
۵۷	..... انتخاب ویژگی
۵۸	..... تبدیل متغیرهای غیر عددی به عددی
۶۵	..... خلق متغیر و ویژگی جدید
۶۶	..... بصری سازی داده ها

## فصل ۴ / یادگیری غیر نظارت شده ..... ۷۰

۷۱	..... مقدمه
۷۳	..... الگوریتم های خوشه بندی
۸۲	..... شناسایی مقادیر دور افتاده
۸۵	..... کاهش ابعاد

**فصل ۵ / یادگیری نظارت شده..... ۸۹**

مقدمه ..... ۹۰

الگوریتم‌های طبقه‌بندی ..... ۹۲

ارزیابی مدل‌های طبقه‌بندی ..... ۱۰۳

الگوریتم‌های رگرسیون ..... ۱۰۵

ارزیابی مدل‌های رگرسیون ..... ۱۰۷

**فصل ۶ / مباحث تکمیلی داده‌کاوی و افزونه‌های نرم‌افزار ORANGE ..... ۱۱۰**

مقدمه ..... ۱۱۱

قوانین انجمنی ..... ۱۱۱

خوشه‌بندی تصاویر ..... ۱۱۵

طبقه‌بندی تصاویر ..... ۱۱۷

تحلیل شبکه‌های اجتماعی ..... ۱۱۹

خوشه‌بندی گراف ..... ۱۲۳

تحلیل سری‌های زمانی ..... ۱۲۴

تحلیل داده‌های جغرافیایی ..... ۱۲۸

کار با داده‌های متنی ..... ۱۳۰

**فصل ۷ / پروژه‌های تکمیلی ..... ۱۳۵**

پیش‌بینی ریزش مشتریان ..... ۱۳۶

پیش‌بینی مقدار هزینه درمان ..... ۱۴۴

خوشه‌بندی جرائم ..... ۱۴۷

پیش‌بینی قیمت سهام در بازار ..... ۱۵۰

خط‌مشی انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب‌هایی با کیفیت عالی است که بتواند  
خواسته‌های به‌روز 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 دست دارید تألیف "جناب آقایان محمد فاتحی، مسعود کاویانی" 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

توسعه هوش مصنوعی از کاربردهای مرسوم به سمت پردازش تصویر و پردازش زبان طبیعی موجب شده است تا نیازهای جدیدی برای این دانش در حوزه های مختلف شکل گیرد. از سوی دیگر با توجه به گسترش هوش مصنوعی و تسهیل کاربرد این دانش در حوزه های مختلف، تالیف کتابی با رویکرد کاربردی و با کمترین میزان کدنویسی ضروری به نظر می رسيد. نرم افزار Orange یک نرم افزار مبتنی بر پایتون است که علاوه بر کمک گرفتن از مزیت ها و قابلیت های پایتون، امکان پیاده سازی پروژه های هوش مصنوعی بدون نیاز به دانش عمیق برنامه نویسی را فراهم ساخته است. افزونه های این نرم افزار نیز امکان پیاده سازی پروژه های پردازش تصویر و پردازش زبان طبیعی را فراهم ساخته است. در این کتاب تلاش شده است انواع کاربردهای مختلف هوش مصنوعی در حوزه های بانکداری، امنیت، صنایع غذایی و... بیان شود. در این کتاب برخی از نوشته ها و مفاهیم که در توضیحات الگوریتم ها و فصل اول آمده اند از کتاب «یادگیری ماشین و یادگیری عمیق با پایتون و R نوشته ی مسعود کاویانی» استخراج شده اند. در پایان نویسندگان این اثر بر خود لازم می دانند تا از زحمات و حمایت های جناب آقای دکتر سعید سعادت مدیریت محترم انتشارات دیباگران تشکر و قدردانی داشته باشند. پایان بخش این مقدمه یک دو بیت از شیخ اشراق است: هان تا سر رشته خرد گم نکنی خود را ز برای نیک و بد گم نکنی رهرو تویی و راه تویی؛ منزل توهشدار که راه خود به خود گم نکنی.

محمد فاتحی

مسعود کاویانی

برای دسترسی به مجموعه های داده و محیط های کاری که در خلال مثال های این کتاب آمده است، می توانید به آدرس [dmml.ir/orange](http://dmml.ir/orange) مراجعه نموده و ابزارها و مثال های گفته شده در این کتاب را دانلود نمایید. همچنین مثال های متعدد دیگری نیز در این آدرس جهت تکمیل یادگیری دانشجویان قرار داده شده است.

ارادتمند شما، مسعود کاویانی

[masoudkaviani.ir](http://masoudkaviani.ir)  
[chistio.ir](http://chistio.ir)