

کلیات داده کاوی

در

علم اطلاعات و دانش شناسی

مؤلفان:

دکتر سعید غفاری

(عضو هیأت علمی دانشگاه پیام نور)

دکتر عمید قاضی زاده

(عضو هیأت علمی دانشگاه پیام نور)





وزارت فرهنگ و ارشاد اسلامی
سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران

سرشناسه	:	غفاری، سعید، ۱۳۴۹ -
عنوان و نام پدیدآور	:	کلیات داده کاوی در علم اطلاعات و دانش‌شناسی / مولفان سعید غفاری، حمید قاضی‌زاده.
مشخصات نشر	:	تهران: اساطیر پارسی: چاپار، ۱۳۹۷.
مشخصات ظاهری	:	۱۸۲ ص: مصور، جدول، نمودار.
شابک	:	978-600-8890-22-5
وضعیت فهرست‌نویسی	:	فیا
یادداشت	:	واژه نامه.
یادداشت	:	کتابنامه.
یادداشت	:	نمونه.
موضوع	:	کتابخانه‌ها -- داده‌پردازی
موضوع	:	Libraries -- Data processing
موضوع	:	داده کاوی
موضوع	:	Data mining
شناسه افزوده	:	قاضی‌زاده، حمید، ۱۳۴۹ -
رده‌بندی کنگره	:	۱۳۹۷ ک۸/غ۷/ز۶۷
رده‌بندی دیویی	:	۰۲۵/۰۰۲۸۵
شماره کتابشناسی ملی	:	۵۱۲۰۴۷۰



تهران: ۸۸۸۹۹۶۸۰ - ۱۲۱۹۶۰۲۱۴

تهران- خیابان ولی عصر (عج) - بالاتر از زرینه رود
کوچه نوریخس - پلاک ۴۰ - نشر چاپار
کلیات داده کاوی در علم اطلاعات و دانش‌شناسی

دکتر سعید غفاری، دکتر حمید قاضی‌زاده

صفحه‌آرا و طراح جلد: رحیم کبیرصابر

ناظر فنی چاپ: مریم تفضلی

چاپ اول ۱۳۹۷

شمارگان: ۵۰۰ نسخه

شابک: ۹۷۸-۶۰۰-۸۸۹۰-۲۲-۵ ISBN: 978-600-8890-22-5

حق هرگونه چاپ و انتشار محفوظ است.

قیمت: ۱۵,۰۰۰ تومان

پیشگفتار

با توجه به تحولات و تغییرات گسترده در عرصه دانش و رشد علوم در عرصه‌های مختلف، این ضرورت احساس شد که متون علمی و دوره‌های آموزشی دانشگاه‌ها و مؤسسات آموزش عالی نیز، مخصوصاً رشته‌های دانشگاهی که به نوعی با موضوع داده و اطلاعات ارتباط تنگاتنگی دارند، دستخوش تغییر و تحول قرار گرفته و در نوع برنامه‌های علمی و واسطه‌ای رسی خود، بازنگری مجددی به عمل آورند. از همین رو چنانچه به‌طور خاص نگاه خود را بر برخی از رشته‌های دانشگاهی معطوف داریم، متوجه خواهیم شد که برای مثال رشته کتابداری در دانشگاه‌ها، هم از ناحیه عنوان رشته و هم از ناحیه سرفصل، با تغییرات کلی همراه بوده است. به گونه‌ای که پیرو مصوبات مربوط به وزارت علوم، تحقیقات و فناوری، عنوان رشته کتابداری به رشته علم اطلاعات و دانش‌شناسی تغییر کرده و همچنین سرفصل آن در مقاطع کارشناسی، کارشناسی ارشد و دکتری نیز دستخوش تغییر و تحول اساسی شده است.

بنابراین با در نظر داشتن تغییرات عمده در سرفصل دروس رشته علم اطلاعات و دانش‌شناسی، عنوان یکی از درس‌ها، به موضوع داده‌کاوی اختصاص یافته است. از همین رو، هرچند یک عنوان کتاب توسط دکتر حمید قاضی‌زاده قبلاً منتشر شده بود، لذا در کتاب پیش رو، مؤلفان تلاش نمودند تا هرچند در کتاب منتشر شده قبلی مطالبی در این حوزه مطرح شده بود، ولی با تمرکز بیشتر روی مباحث مطروحه، کتاب حاضر را مطابق سرفصل وزارت متبوع تألیف کرده و در اختیار خوانندگان محترم قرار دهند. به زعم نظر مؤلفان کتاب پیش رو، اگر ساختار داده‌کاوی را به دو لایه پوسته و هسته

تقسیم‌بندی کنیم، لایه پوسته بر جنبه تعاریف و مفاهیم متمرکز شده و لایه هسته بیشتر بر تحلیل‌های آماری، فرمول‌ها و معادلات ریاضی نهاده شده است. لذا در این کتاب با توجه به اینکه بیشتر محققان علوم انسانی مخصوصاً رشته علم اطلاعات و دانش‌شناسی با مباحث عمیق و فرمول‌های بنیادین ریاضیات، فاصله داشته و کمتر در این حوزه ورود کرده‌اند، بنابراین صلاح دیده شد تا بیشتر محتویات کتاب پیش رو، بر جنبه پوسته یا همان مفاهیم و تعاریف معطوف شده و سعی شد تا جز مواردی که ضرورت ایجاب می‌نماید، از ذکر تحلیل‌ها و فرمول‌های آماری پرهیز شود. امید است تا با اتکال بر خداوند، منان، ساختار هسته یا محتوای آماری و ریاضی مباحث مرتبط با داده‌کاوی به‌طور مبسوط در کتاب یگری مطرح شود.

همچنین نبود شعر و کتاب بیشتر در این زمینه و مخصوصاً به زبان فارسی برای دانشجویان در مقاطع تحصیلی مختلف، زمینه‌ای شد تا مؤلفان برای این موضوع تأملی اندیشیده و کتاب حاضر با عنوان کلیات داده‌کاوی در علم اطلاعات و دانش‌شناسی را به رشته تحریر درآوردند. شایان ذکر است یادآور شود که مخاطبان این کتاب نه تنها دانشجویان و محققان حوزه کتابداری، اطلاع‌رسانی و علم اطلاعات می‌باشند، بلکه سایر دانشجویان و محققان رشته‌های علوم انسانی که قصد داشته تا چشم‌انداز مشخصی از داده‌کاوی به‌دست آورده و یا با مفاهیم و تعاریف اولیه مرتبط با داده‌کاوی آشنا شوند نیز، می‌توانند از جمله مخاطبان کتاب حاضر باشند.

با عنایت به اینکه در سرفصل مصوب وزارت علوم و تحصیلات فراوری، برای درس داده‌کاوی مباحثی از قبیل: مفاهیم و تعاریف داده‌کاوی، اکتشاف دانش، داده‌کاوی برای بازاریابی، الگوریتم‌های داده‌کاوی، پاکسازی داده‌ها، یادگیری ماشین، روش‌های انبارش داده، پردازش پیوسته داده‌ها، خوشه‌بندی، وب‌کاوی و آینده داده‌کاوی مطرح شده است، مؤلفان کتاب پیش رو نیز تلاش نمودند تا مطابق با مباحث مطروحه سرفصل وزارت متبوع، کتاب حاضر را در چهار فصل با موضوعات مبانی اولیه داده‌کاوی، ساختار داده‌کاوی، داده‌کاوی و عصر دیجیتالی و داده‌کاوی و کتابخانه به رشته تألیف درآورند.

در انتها شایسته است یادآوری نماید که کتاب کلیات داده‌کاوی در علم اطلاعات و دانش‌شناسی به‌منظور آشنایی بیشتر اساتید و دانشجویان رشته علم اطلاعات و دانش‌شناسی، پرکردن خلاء درسی موجود در این رشته و تقریباً به‌مثابه طلیعه‌ای جهت تدوین کتب با موضوع داده‌کاوی بوده و با این هدف تألیف شد تا فتح بابی باشد برای چاپ کتب جدید داده‌کاوی به همراه محاسبات دقیق آماری، عملی و کاربردی در حوزه کتابداری، علم اطلاعات و دانش‌شناسی. بی‌شک این کتاب حال از نقص و ایراد نبوده و برای بهبود و لحاظ‌کردن اندیشه‌های بدیع خوانندگان و صاحب‌نظران در مراحل بعدی، نیازمند دریافت نظرها و پیشنهادهای عالمانه صاحبان فن در حوزه علم اطلاعات و کتابداری می‌باشد تا با ارائه دیدگاه‌های اندیشمندانه خود، مؤلفان را در ویرایش‌های بعدی کتاب، یاری رسانند.

دکتر سعید غفاری و دکتر حمید قاضی‌زاده

زمستان ۱۳۹۶

فهرست مطالب

پیشگفتار	۵
فصل اول مبانی اولیه داده کاوی	۱۳
۱. مقدمه	۱۵
۲. تاریخچه داده کاوی	۱۶
۳. داده، اطلاعات و دانش	۱۹
۴. داده کاوی چیست و چرا به آن نیاز داریم؟	۲۰
۵. چرا داده کاوی؟	۲۱
۶. الگوریتم‌ها و فنون داده کاوی	۲۳
۷. فنون داده کاوی	۲۳
۸. نقش داده کاوی در تولید دانش	۲۵
۹. جایگاه داده کاوی در سازمان	۲۵
۱۰. داده کاوی و دیگر رشته‌ها	۲۶
۱۱. آمار و داده کاوی	۲۷
۱۲. نام‌های دیگر داده کاوی	۲۸
۱۳. تعریف داده کاوی	۲۹
۱۴. دو دیدگاه در مورد داده کاوی	۳۰
۱۵. آماده‌سازی و پیش پردازش داده‌ها	۳۱
۱۶. ساختار یک سیستم داده کاوی	۳۴
۱۷. معیارهای جذابیت الگو	۳۵
۱۸. ارزیابی و به تصویر کشیدن الگوهای کشف‌شده	۳۶
۱۹. انواع داده کاوی	۳۶
۲۰. وظایف داده کاوی	۳۷
۲۰-۱. کاربردهای داده کاوی	۴۲
۲۰-۲. محدودیت‌های داده کاوی	۴۷

۴۸.....	۲۰-۳. مشکلات داده‌کاوی.....
۴۹.....	۲۰-۴. نمونه‌های اجرا شده داده‌کاوی.....
۴۹.....	۲۰-۵. تعامل داده‌کاوی و سازمان.....
۵۲.....	۲۰-۶. داده‌کاوی در بازاریابی.....
۵۹.....	فصل دوم: ساختار داده‌کاوی.....
۶۱.....	۲۱. پردازش تراکنش آنلاین و پردازش تحلیلی برخط (OLAP و OLTP).....
۶۱.....	۲۱-۱. پردازش تحلیلی برخط چیست؟.....
۶۲.....	۲۱-۲. نحوه عملکرد سیستم‌های OLAP.....
۶۴.....	۲۱-۳. انواع مختلف سیستم‌های پردازش تحلیلی برخط.....
۶۴.....	۲۲. جاریه، نیاز داده.....
۶۵.....	۲۲-۱. چهار خصوصیت اصلی ابزار داده‌ها.....
۶۶.....	۲۲-۲. تفاوت داده و ابزار داده.....
۶۷.....	۲۲-۳. زیای و معایب پایگاه داده.....
۶۸.....	۲۳. الگوریتم‌های داده‌کاوی.....
۶۸.....	۲۳-۱. شبکه‌های عصبی.....
۷۱.....	۲۳-۲. قابلیت‌های شبکه‌های عصبی.....
۷۱.....	۲۳-۳. الگوریتم درخت تصمیم.....
۷۳.....	۲۳-۴. الگوریتم قواعد پیوند.....
۷۳.....	۲۳-۵. مثال‌هایی از قواعد پیوند.....
۷۴.....	۲۳-۶. الگوریتم‌های زنجیره.....
۷۵.....	۲۳-۷. خوشه‌بندی (Clustering).....
۷۷.....	۲۳-۸. هدف از خوشه‌بندی.....
۷۸.....	۲۳-۹. تفاوت خوشه‌بندی فازی داده و کلاسیک.....
۸۲.....	۲۳-۱۰. خوشه‌بندی سلسله‌مراتبی.....
۸۳.....	۲۳-۱۱. معیار خوشه‌بندی.....
۸۵.....	فصل سوم: داده‌کاوی و عصر دیجیتال.....
۸۷.....	۲۴. نرم‌افزارهای داده‌کاوی.....
۹۱.....	۲۵. وب‌کاوی.....
۹۳.....	۲۵-۱. کاوش داده روی وب.....
۹۴.....	۲۵-۲. داده‌کاوی در وب روی پایگاه داده‌ای رابطه‌ای Web.....
۹۴.....	۲۶. داده‌کاوی چند رسانه‌ای.....
۹۵.....	۲۷. داده‌کاوی و بصرسازی در اینترنت.....
۹۶.....	۲۸. متن‌کاوی و کاربرد آن در سازمان‌دهی دانش.....
۹۹.....	۲۹. مراحل سه‌گانه متن‌کاوی.....
۱۰۰.....	۲۹-۱. استفاده از فنون متن‌کاوی در سازمان‌دهی دانش.....

۱۰۵	۲۹-۲. کاربردهای متن کاوی
۱۰۸	۲۹-۳. فرایند متن کاوی
۱۰۹	۲۹-۴. روش های متن کاوی
۱۱۰	۲۹-۵. چرخه داده کاوی نوین
۱۱۱	۳۰. واحدهای حافظه رایانه
۱۱۴	۳۱. کلان داده (بیگ دیتا)
۱۱۵	۳۱-۱. چالش ها و خصوصیات کلان داده
۱۱۷	۳۱-۲. ابزارهای ذخیره و پردازش در حوزه کلان داده
۱۱۸	۳۱-۳. کاربردهای کلان داده
۱۱۹	۳۱-۴. رهیافت اصلی کلان داده
۱۲۱	۳۲. بانک اطلاعاتی رابطه ای
۱۲۵	۳۳. بانک اطلاعات غیررابطه ای
۱۲۶	۳۴. پایگاه داده های غیر رابطه ای NOSQL
۱۳۰	۳۵. ارتباط میان داده کاوی و پایگاه داده ها
۱۳۲	۳۶. انواع داده
۱۳۳	۳۶-۱. انواع داده در سازمان
۱۳۶	۳۶-۲. تعریف داده های سازمانی
۱۳۷	۳۶-۳. گسترش داده های ساخت نیافت
۱۳۸	۳۷. روش های تحلیل داده ها
۱۴۰	۳۷-۱. روش ان. ال. بی. NLP
۱۴۲	۳۷-۲. روش اس. وی. ام. SVM
۱۴۳	۳۸. خلاصه سازی متن
۱۴۸	۳۹. شبکه وازگان فارسی، فارس نت
۱۵۱	فصل چهارم: داده کاوی و کتابخانه
۱۵۳	۴۰. داده کاوی در کتابخانه ها
۱۵۵	۴۱. بهره گیری از مفهوم داده کاوی در کتابخانه
۱۵۶	۴۲. کاربردهای داده کاوی در کتابخانه ها
۱۶۲	۴۳. کاربردهای بالقوه داده کاوی در کتابخانه ها
۱۷۱	۴۴. آینده داده کاوی در کتابخانه ها
۱۷۵	منابع
۱۷۹	نمایه