

۹۸۷۶۵۴
به نام خدا

یادگیری ماشین و علم داده

مبانی، مفاهیم، الگوریتم‌ها و ابزارها

www.ketab.ir

تالیف و گردآوری: میلاد وزان

سرشناسه	: روان، میلاد، ۱۳۷۱-
عنوان و نام پدیدآور	: یادگیری ماشین و علم داده: مبانی، مفاهیم، الگوریتمها و ابزارها/ تالیف و گردآوری: میلاد وزان.
مشخصات نشر	: تهران: میعاد اندیشه، ۱۴۰۰.
مشخصات ظاهری	: ۲۸۷ ص: مصور، جدول، نمودار.
شابک	: 978-622-231-851-2
وضعیت فهرست نویسی	: فیبا
یادداشت	: کتابنامه.
موضوع	: فراگیری ماشینی Machine learning پایتون (زبان برنامه نویسی کامپیوتر) Python (Computer program language) یادگیری عمیق Deep Learning ساختار داده‌ها -- الگوهای ریاضی Data structures (computer science) -- Mathematical models
رده بندی کنگره	: Q۳۲۵/۵
رده بندی دیویی	: ۰۰۶/۳۱
شماره کتابشناسی ملی	: A۷۲۷۶۱۹
اطلاعات رکورد کتابشناسی	: فیبا



Miaadpub.ir

تلفن انتشارات: ۰۹۱۲۵۱۲۰۰۶۷ - ۰۲۱۶۶۰۱۷۴۴۸ - ۰۲۱۶۶۰۱۴۷۹۷

عنوان کتاب: یادگیری ماشین و علم داده: مبانی، مفاهیم، الگوریتمها و ابزارها
تالیف و گردآوری: میلاد وزان
ناشر: میعاد اندیشه
نوبت چاپ: اول - ۱۴۰۰
شمارگان: ۱۰۰۰ نسخه
قیمت: ۱/۲۰۰/۰۰۰ ریال
شابک: ۹۷۸ - ۶۲۲ - ۲۳۱ - ۸۵۱ - ۲

همه ی حقوق مادی و معنوی این اثر برای مولف محفوظ است.
تکثیر و انتشار این اثر به هر صورت، از جمله بازنویسی، فتوکپی، ضبط الکترونیکی
و ذخیره در سیستم های بازیابی و پخش، بدون دریافت مجوز کتبی و قبلی از مولف
به هر شکلی ممنوع است.
این اثر تحت حمایت «قانون حمایت از حقوق مولفان، مصنفان و هنرمندان ایران» قرار دارد

پیش‌گفتار

علم داده و یادگیری ماشین

امروزه اگر برجسته‌ترین رهبران تجارت جهانی را گرد هم بیاورید و از آنان بخواهید که بزرگ‌ترین تفاوت بین کسب‌وکار در قرن بیستم در مقابل قرن بیست‌ویکم را بیان کنند، با احتمال خیلی زیاد، آن‌ها یک کلمه را خواهند گفت: داده.

از آغاز قرن، مقدار داده‌ها با ظهور رسانه‌های اجتماعی، تلفن‌های هوشمند، اینترنت اشیا و دیگر پیشرفت‌های فن‌آوری با سرعت شگفت‌انگیزی افزایش یافته است. تخمین زده می‌شود که بیش از ۹۰ درصد از کل داده‌های ایجاد شده توسط انسان‌ها در پنج سال گذشته تولید شده‌اند. این انفجار اطلاعات به عنوان "کلان داده" شناخته می‌شود و به طور کامل جهان اطراف ما را متحول خواهد کرد. رشد نمایی در تولید داده‌ها، سازمان‌ها را بیش از هر اندازه‌ای به این فکر فروبرده است که چگونه می‌توانند از داده‌ها برای تحقق منافع تجاری خود استفاده کنند. در همین حال، افراد به طور فزاینده‌ای به دنبال توسعه مهارت‌های داده خود برای برجسته کردن رزومه، پیشرفت شغلی و کسب امنیت شغلی هستند.

امروزه، داده‌ها به عنوان ابزار و سوختی برای کسب و کارها در جهت بدست آوردن بینش‌های مهم و بهبود عملکرد آن‌ها هستند. علم داده‌ها تقریباً تمام صنایع جهان را تحت سلطه خود درآورده است. امروزه هیچ صنعتی در دنیا وجود ندارد که از داده‌ها استفاده نکند. اما چه کسی این بینش را بدست خواهد آورد؟ چه کسی تمام داده‌های خام را پردازش می‌کند؟ همه چیز توسط یک تحلیل‌گر داده یا یک دانشمند داده انجام می‌شود. این‌ها محبوب‌ترین نقش شغلی در این زمینه هستند. چرا که شرکت‌ها در سراسر جهان به دنبال بیشترین استفاده از داده‌ها هستند. برای افرادی که به دنبال پتانسیل شغلی بلند مدت هستند، مشاغل علم داده از مدت‌هاست که گزینه‌ی مطلوب و اطمینان‌بخشی هستند. این جریان با ادغام هوش مصنوعی و یادگیری ماشین در زندگی روزمره ما و اقتصاد، احتمالاً ادامه‌دار خواهد بود.

داده‌ها فقط در مورد گذشته به ما نمی‌گویند. اگر آن را با دقت و با روش‌های دقیق مدل‌سازی کنیم، می‌توانیم الگوها و همبستگی‌هایی را برای پیش‌بینی بازارهای سهام، تولید توالی پروتئین، کشف ساختارهای بیولوژیکی همانند ویروس‌ها و موارد دیگر پیدا کنیم. با این حال، مدل‌سازی حجم زیادی از داده‌ها به صورت دستی کار خسته‌کننده‌ای است. برای رسیدن به این هدف، به الگوریتم‌های یادگیری ماشین روی آورده‌ایم که می‌توانند به ما در استخراج اطلاعات از داده‌ها کمک کنند. یادگیری ماشین خودکارسازی وظایفی که معمولاً به هوش انسانی نیاز دارند را انجام می‌دهند. به نوبه خود، یادگیری ماشین را می‌توان به عنوان استفاده و توسعه سیستم‌های رایانه‌ای تعریف کرد که قادر به یادگیری و تطبیق بدون برنامه‌ریزی صریح هستند. این سیستم‌ها از الگوریتم‌ها و مدل‌های آماری برای تحلیل و استنتاج از الگوهای موجود در داده‌ها استفاده

می‌کنند. الگوریتم‌های یادگیری ماشین می‌توانند مسائل پیچیده‌ای را که انجام دستی آن‌ها غیرعملی و یا حتی غیرممکن است را حل کنند و توزیع‌ها، الگوها و همبستگی‌ها را بیاموزند تا دانش درون داده‌ها را آشکار کنند. الگوریتم‌ها این کار را با کاوش یک مجموعه داده و ایجاد یک مدل تقریبی بر روی توزیع داده‌ها انجام می‌دهند، به طوری که وقتی داده‌های جدید و دیده نشده را تغذیه می‌کنیم، نتایج خوبی به همراه خواهد داشت.

به طور کلی، الگوریتم‌های یادگیری ماشین می‌توانند با سه رویکرد متفاوت توانایی یادگیری بدست آورند:

- **یادگیری بانظارت:** یادگیری بانظارت در یادگیری ماشین روشی است که مدل را برای پیش‌بینی نتیجه بر اساس داده‌های برچسب‌گذاری شده ایجاد می‌کند. وجود داده‌های برچسب‌دار به این معنا است که برای هر نمونه‌ای از مجموعه داده، یک پاسخ یا راه‌حل داده شده است. به عنوان یک مثال ساده، اگر بخواهیم مدل یادگیری ماشین ما پیش‌بینی در مورد میوه‌های سیب یا موز داشته باشد، برچسب مقادیر "سیب" یا "موز" را به همواره مجموعه‌ای از ویژگی‌ها همانند وزن، طول، عرض و هر اندازه‌گیری مرتبط از میوه‌های موجود را می‌گیرد. بیا بیا به یک مثال مرتبط‌تر با کسب و کار نگاه کنیم؛ ریزش مشتری برای یک شرکت بهتر ریزش مشتری، ابتدا باید تجزیه و تحلیل کنید که چه شاخص‌هایی ممکن است منجر به خروج مشتری شود. مجموعه داده شما برای این نوع مدل شامل متغیرهای داخلی همانند، روزهای سپری شده پس از آخرین خرید، میانگین مقدار خرید و همچنین متغیر پیش‌بینی‌کننده برچسب‌گذاری شده، یعنی این که آیا فرد هنوز مشتری است یا خیر. از آنجایی که ما داده‌های گذشته در مورد وضعیت مشتری را داریم، ایجاد مدلی با این نوع مجموعه داده می‌تواند کاندیدای عالی برای یادگیری بانظارت باشد.

- **یادگیری غیرنظارتی:** یادگیری غیرنظارتی در یادگیری ماشین زمانی است که نمونه‌ها را بدون هیچ راهنمایی به الگوریتم ارائه می‌کنیم و ایجاد برچسب را به الگوریتم واگذار می‌کنیم. به عبارت دیگر، یادگیری غیرنظارتی، به یافتن الگوهای پنهان از داده‌های بدون برچسب و ایجاد گروه‌ها و خوشه‌ها می‌پردازد. به عنوان مثال، زمانی که ما باید بفهمیم که چگونه گروه‌های داخل مجموعه داده مشتری را می‌توان بر اساس ویژگی‌ها و رفتارهایشان به بخش‌های مشابه دسته‌بندی کرد. یادگیری غیرنظارتی اغلب برای تجزیه و تحلیل اکتشافی و تشخیص ناهنجاری استفاده می‌شود. چراکه به یافتن نحوه ارتباط بخش‌های داده و اینکه چه روندهایی ممکن است وجود داشته باشد، کمک می‌کند. آن‌ها می‌توانند برای پیش‌پردازش داده‌های شما قبل از استفاده از الگوریتم یادگیری بانظارت استفاده شوند.

• **یادگیری تقویتی:** یادگیری تقویتی تکنیکی است که بازخورد آموزشی را با استفاده از مکانیزم پاداش ارائه می‌کند. فرآیند یادگیری به عنوان یک عامل رخ می‌دهد که با یک محیط تعامل دارد و روش‌های مختلف را برای رسیدن به یک نتیجه امتحان می‌کند. عامل هنگامی که به وضعیت مطلوب یا نامطلوب برسد پاداش یا تنبیه دریافت می‌کند. از طریق این بازخورد یادگیری، عامل یاد می‌گیرد که کدام حالت‌ها منجر به نتایج خوب و کدام منجر به شکست می‌شود و باید از آن‌ها اجتناب کرد. به عنوان مثال، زمانی که برای عملکرد موفقیت‌آمیز در یک محیط رقابتی، مانند یک بازی ویدیویی یا بازار سهام، به نرم‌افزار نیاز داریم، می‌توانیم از یادگیری تقویتی استفاده کنیم. در این حالت، نرم‌افزار شروع به فعالیت در محیط می‌کند و مستقیماً از خطاهای خود درس می‌گیرد تا زمانی که مجموعه‌ای از قوانینی را پیدا کند که موفقیت آن را تضمین می‌کند. یادگیری تقویتی به داده‌های برچسب‌دار، همانند یادگیری بانظارت نیاز ندارد. علاوه بر این، حتی از یک مجموعه داده بدون برچسب، همانند یادگیری غیرنظارتی استفاده نمی‌کند. یادگیری تقویتی به جای تلاش برای کشف رابطه در یک مجموعه داده، به طور مداوم در بین نتایج تجربیات گذشته خود و همچنین ایجاد تجربیات جدید بهینه می‌شود. به عبارت دیگر، مجموعه داده‌ها و نتایج جدیدی را با هر تلاش ایجاد می‌کند.

یادگیری بانظارت در علم داده بسیار مهم است، چراکه به ما اجازه می‌دهد تا چیزی را که نسل بشر آرزوی آن را داشته است، انجام دهیم: پیش‌بینی. پیش‌بینی در تجارت و برای سودمندی کاربرد خیلی زیادی دارد و ما را قادر می‌سازد تا بهترین اقدام را انجام دهیم، چراکه از طریق پیش‌بینی، نتیجه احتمالی یک موقعیت را می‌دانیم.

یادگیری بانظارت ممکن است برای برخی همانند جادوگری به نظر برسد. با این حال، یادگیری بانظارت به هیچ وجه جادو نمی‌کند. بلکه، یادگیری بانظارت بر اساس دستاوردهای انسان در ریاضیات و آمار و با استفاده از تجربیات و مشاهدات انسانی و تبدیل آن‌ها به پیش‌بینی‌های دقیق به گونه‌ای که ذهن هیچ انسانی قادر به انجام آن نیست، کمک می‌کند.

با این حال، یادگیری بانظارت تنها در شرایط مطلوب خاصی می‌تواند پیش‌بینی کند. از این رو، برای این کار، بسیار مهم است که نمونه‌هایی از گذشته داشته باشیم که می‌توانیم از آن‌ها قوانین و نکاتی استخراج کنیم که می‌تواند از جمع‌بندی آن‌ها یک پیش‌بینی بسیار محتمل ایجاد شود.

چرا یادگیری ماشین؟

حجم داده‌های در دسترس ما به طور مداوم در حال افزایش است. ماشین‌ها از این داده‌ها برای یادگیری و بهبود نتایج و ارائه آن‌ها به ما استفاده می‌کنند. این نتایج می‌تواند در ارائه پیش‌بینی‌های ارزشمند و همچنین اتخاذ تصمیمات تجاری آگاهانه بسیار مفید باشد. یادگیری ماشین به طور

مداوم در حال رشد است و در پی آن، کاربردهای یادگیری ماشین نیز رو به رشد هستند. ما بیش از آنچه می‌دانیم از یادگیری ماشین در زندگی روزمره خود استفاده می‌کنیم. یادگیری ماشین خود را وارد زندگی روزمره ما کرده است، حتی بدون اینکه ما متوجه شویم. الگوریتم‌های یادگیری ماشین به دنیای اطراف ما قدرت داده‌اند. ناگفته نماند که آینده از قبل در اینجاست و یادگیری ماشین نقش مهمی در نحوه پنداره‌های معاصر ما دارد.

امروزه یادگیری ماشین تمام توجهی که به آن نیاز دارد را دارد. یادگیری ماشین می‌تواند بسیاری از وظایف را به‌طور خودکار انجام دهد، به خصوص آن‌هایی که تنها انسان‌ها می‌توانند با هوش ذاتی خود انجام دهند. تکثیر این هوش در ماشین‌ها تنها با کمک یادگیری ماشین قابل دستیابی است.

با کمک یادگیری ماشین، کسب‌وکارها می‌توانند کارهای روتین را خودکار کنند. همچنین به خودکارسازی و ایجاد مدل‌هایی برای تجزیه و تحلیل داده‌ها کمک می‌کند. صنایع مختلف برای بهینه‌سازی عملکرد خود و اتخاذ تصمیمات هوشمندانه به مقادیر زیادی از داده‌ها وابسته هستند. یادگیری ماشین به ایجاد مدل‌هایی کمک می‌کند تا بتوانند مقادیر زیادی از داده‌های پیچیده را پردازش و تجزیه و تحلیل کنند و نتایج دقیقی را ارائه کنند. این مدل‌ها، دقیق و مقیاس‌پذیر هستند و با تابع زمانی کم‌تری کار می‌کنند. با ساخت چنین مدل‌های دقیق یادگیری ماشین، کسب‌وکارها می‌توانند از فرصت‌های سودآور استفاده کنند و از ریسک‌های ناشناخته اجتناب کنند.

تشخیص تصویر، تولید متن، طبقه‌بندی متن، تشخیص بیماری‌ها و بسیاری از موارد دیگر در دنیای واقعی کاربرد دارند. از این‌رو، این امر زمینه را برای درخشش کارشناسان یادگیری ماشین به عنوان یک متخصص حرفه‌ای افزایش می‌دهد. علاوه بر این، با توجه به سرعت سریعی که در جهش‌های فناوری انجام شده است، بسیاری از شرکت‌ها از فناوری عقب مانده‌اند. تحول دیجیتال صنعت بزرگی است و حقیقت موضوع این است که به اندازه کافی متخصص یادگیری ماشین برای پاسخگویی به نیازهای صنعت جدید وجود ندارد.

اگر می‌خواهید حرفه‌ی خود را به سطح دیگری ببرید، یادگیری ماشین می‌تواند این کار را برای شما انجام دهد. اگر به دنبال این هستید که خود را درگیر کاری کنید که شما را بخشی از چیزی کند که هم جهانی و هم مرتبط با معاصر باشد، یادگیری ماشین می‌تواند این کار را برای شما انجام دهد.

ارتباط یادگیری ماشین و علم داده

یادگیری ماشین، تنها در صورتی می‌تواند بینش‌های ارزشمندی ارائه دهد که داده‌های با کیفیت را دریافت کند. از این‌رو، بدون استفاده از داده‌های تمیز، سازگار و با کیفیت، بینش معنی‌دار کمی (در صورت وجود) می‌توان تولید کرد. در عین حال، دانشمند داده به یادگیری ماشین نیاز

دارد، زیرا درک و پیش‌بینی دقیق نتایج حاصل از حجم عظیمی از داده‌های پیچیده که سازمان‌ها در اختیار دارند، عملاً غیرممکن است.

دانشمندان داده همچنین باید حس تجاری عالی داشته باشند تا بفهمند چه چیزی باعث موفقیت کسب‌وکار می‌شود، کجا می‌توان آن را بهبود بخشید و گزینه‌های احتمالی برای دستیابی به آن نتیجه چیست. یادگیری ماشین را می‌توان آمار کاربردی برای انجام این امر در نظر گرفت. به عبارت دیگر، یادگیری ماشین، ادغام علوم رایانه و رشته‌های مختلف ریاضی است که در آن از مفاهیم علوم رایانه برای ساختن مدل‌های ریاضی قوی استفاده می‌شود که می‌تواند مجموعه‌ای از مسائل مشابه و مرتبط را حل کند.

دانشمندان داده نگران به اطمینان از این‌که آیا مدل یادگیری ماشین به اهداف پروژه دست می‌یابد، هستند. اینجاست که مجموعه مهارت‌های تجاری شاید مهم‌ترین مهارت برای داشتن باشد. برای موفقیت در توسعه مدل یادگیری ماشین، دانشمندان داده باید درک معقولی از مشکل در دست و اهداف پروژه داشته باشند. بدون این مورد، احتمال موفقیت کمی برای هر برنامه علم داده و مدل یادگیری ماشین وجود دارد. در این راستا، تقریباً ۸۰ درصد از زمان یک دانشمند داده صرف کاوش، تمیزسازی و آماده‌سازی داده‌ها می‌شود. انجام درست این کار بخشی اساسی از فرآیند است. پس از تکمیل، دانشمند داده می‌تواند توسعه مدل یادگیری ماشین را شروع کند. آنها می‌توانند مدل‌های مختلف را آزمایش و مقایسه کنند و سپس امیدوارکننده‌ترین نامزد را برای عرضه در محیط تولید، بهینه‌سازی کنند.

یکی از موثرترین راه‌ها برای مشاهده‌ی پنهانی این مدل‌ها از طریق مصورسازی داده‌ها است. ارائه یک گزارش با داده‌ها، رهبران کسب وکار را قادر می‌سازد تا تصمیمات آگاهانه‌تری بگیرند که می‌تواند به نفع سازمان باشد. جدای از تهیه داده‌ها، این مصورسازی شاید مهم‌ترین گام در کمک به اطمینان از موفقیت پروژه باشد.

بنابراین، با وجود اینکه مدل‌های یادگیری ماشین مهم هستند، موفقیت آن‌ها به شدت وابسته به توانایی تیم داده‌ها برای درک و فراهم کردن داده‌های ساختار یافته با اطلاعات عالی است که به مدل اجازه می‌دهد تا پیش‌بینی‌های دقیق انجام دهد.

علم داده و یادگیری ماشین به یکدیگر وابسته هستند و برای موفقیت یک سازمان داده‌محور بسیار مهم هستند. با این حال، همه چیز به کیفیت داده‌های مورد استفاده بستگی دارد.

درباره‌ی کتاب

کتاب حاضر شامل دو بخش مقدمات و یادگیری ماشین است که در مجموع شامل ۹ فصل می‌شود. عمده‌ی مطالب اصلی و مهم کتاب در بخش دوم می‌باشد. از این‌رو، اگر با مفاهیم برنامه‌نویسی و پیش‌پردازش داده‌ها آشنایی دارید، بخش اول کتاب مناسب شما نیست و می‌توانید از این بخش عبور کنید و مستقیماً وارد بخش یادگیری ماشین شوید.

فصل ۱: علم داده

۲۴.....	علم داده چیست؟
۲۵.....	کلان داده چیست؟
۲۶.....	تفاوت علم داده با کلان داده
۲۷.....	چرا علم داده را یاد بگیریم؟
۲۷.....	تجزیه و تحلیل داده
۲۹.....	مسئولیت‌های اصلی یک تحلیل‌گر داده
۲۹.....	انواع تجزیه و تحلیل داده
۳۰.....	تجزیه و تحلیل توصیفی - چه اتفاقی افتاده است
۳۰.....	تجزیه و تحلیل تشخیصی - چرا این اتفاق افتاد
۳۱.....	تجزیه و تحلیل پیش‌گویانه - چه اتفاقی خواهد افتاد
۳۱.....	تجزیه و تحلیل تجویزی - چه اقدامی باید انجام شود
۳۳.....	چرخه دوام علم داده
۳۴.....	دانشمند داده
۳۵.....	مهندس داده
۳۵.....	نقش مهندس داده
۳۶.....	حوزه‌ها و مهارت‌های اساسی مطالعه در علم داده
۳۶.....	یادگیری ماشین
۳۶.....	یادگیری عمیق
۳۷.....	ریاضیات
۳۷.....	آمار و احتمالات
۳۷.....	پردازش زبان طبیعی
۳۷.....	مصورسازی داده‌ها
۳۷.....	زبان برنامه‌نویسی
۳۷.....	الگوریتم
۳۸.....	کاربرد علم داده
۳۸.....	حمل و نقل
۳۸.....	تشخیص رسیک و کلاه‌برداری
۳۸.....	ژنتیک و ژنومیکس
۳۸.....	توسعه دارو

۳۹	 خلاصه فصل
۳۹	 مراجع برای مطالعه بیشتر

فصل ۲: مقدمه‌ای بر پایتون

۴۲	 پایتون چیست؟
۴۲	 دلیل انتخاب پایتون در علم داده؟
۴۳	 نصب پایتون
۴۳	 بارگیری
۴۳	 نصب
۴۴	 شروع کار با پایتون
۴۵	 کتابخانه‌ها و مدیریت آن‌ها در پایتون
۴۸	 نصب، حذف و بروزرسانی کتابخانه‌ها
۵۰	 جویپتر نوت‌بوک
۵۰	 نصب جویپتر
۵۰	 اجرای جویپتر و ایجاد یک نوت‌بوک جدید
۵۲	 کار در جویپتر نوت‌بوک
۵۳	 برنامه‌نویسی در پایتون
۵۳	 نحو در زبان پایتون
۵۴	 کلمات کلیدی
۵۴	 شناسه
۵۵	 متغیر
۵۶	 انواع داده
۵۶	 عددی
۵۶	 رشته‌ها
۵۷	 عملگرها
۵۷	 عملگرهای حسابی
۵۸	 عملگرهای منطقی
۵۸	 عملگرهای مقایسه‌ای
۵۹	 داده ساختارها
۵۹	 لیست
۶۳	 تاپل
۶۴	 دیکشنری
۶۶	 مجموعه

۶۷.....	ساختارهای کنترلی و حلقه‌ها
۶۷.....	دستورات شرطی
۷۰.....	حلقه‌ها
۷۱.....	توابع
۷۲.....	تعریف تابع
۷۳.....	کار با کتابخانه NumPy
۷۴.....	وارد کردن NumPy
۷۴.....	ایجاد یک آرایه NumPy
۷۵.....	آرایه‌ای از صفرها
۷۶.....	آرایه‌ای از یک‌ها
۷۶.....	افزودن، حذف و مرتب‌سازی عناصر
۷۷.....	شناسایی شکل و اندازه یک آرایه
۷۷.....	خلاصه فصل
۷۸.....	مراجع برای مطالعه بیشتر

فصل ۳: داده

۸۰.....	داده
۸۰.....	نوع داده کیفی
۸۱.....	اسمی
۸۱.....	ترتیبی
۸۲.....	نوع داده کمی
۸۲.....	گسسته
۸۲.....	پیوسته
۸۳.....	داده‌های سری زمانی
۸۴.....	مجموعه داده‌ها و ویژگی‌ها
۸۵.....	ویژگی‌ها کلی مجموعه داده‌ها
۸۶.....	نمونه‌هایی از داده‌های با ابعاد بالا
۸۶.....	نحوه مدیریت داده‌های با ابعاد بالا
۸۸.....	گردآوری و ساخت داده
۸۸.....	وب تراش
۸۹.....	تکنیک‌های وب تراش
۹۲.....	کتابخانه‌های وب تراش
۹۶.....	ذخیره‌سازی و ارائه داده‌ها

۹۶	 CSV (مقادیر جداشده با ویرگول)
۹۶	 نحوه وارد کردن داده‌های CSV در پایتون
۹۹	 XML (زبان نشانه‌گذاری گسترش‌پذیر)
۱۰۰	 نحوه وارد کردن داده‌های XML در پایتون
۱۰۱	 JSON (نشانه‌گذاری شیء جاوااسکریپت)
۱۰۲	 نحوه وارد کردن داده‌های JSON در پایتون
۱۰۴	 پیش‌پردازش و آماده‌سازی داده‌ها
۱۰۴	 تمیزسازی داده‌ها
۱۰۵	 مقادیر مفقود شده
۱۰۷	 داده‌های دورافتاده
۱۱۰	 داده‌های تکراری
۱۱۰	 داده‌های بی‌ربط
۱۱۰	 تبدیل داده‌ها
۱۱۱	 تجمیع
۱۱۱	 گسسته‌سازی
۱۱۲	 هموارسازی
۱۱۲	 مقیاس‌بندی
۱۱۴	 متعارف‌سازی
۱۱۶	 هنجارسازی
۱۱۶	 متعارف‌سازی یا هنجارسازی؟
۱۱۷	 مصورسازی داده‌ها
۱۱۸	 اهمیت و مزایای مصورسازی داده‌ها
۱۱۹	 هدف از مصورسازی داده‌ها چیست؟
۱۱۹	 از چه نوع نمودار مصورسازی استفاده کنیم؟
۱۲۰	 انواع نمودارهای مصورسازی داده‌ها
۱۲۱	 نمودار خطی
۱۲۶	 نمودار میله‌ای
۱۲۷	 نمودار ستونی
۱۳۱	 نمودار دایره‌ای
۱۳۳	 نمودار نقطه‌ای
۱۳۶	 نمودار مساحت
۱۳۸	 نمودار حبابی
۱۴۰	 خلاصه فصل