



سازمان جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی



دوره داده کاوی با پایتون

داده کاوی (Data Mining) علم استخراج الگوها، اطلاعات و تحلیل از مجموعه داده‌های خامی است که در یک سازمان و یا یک جامعه یا هر مجموعه دیگری تولید شده است. به مجموعه‌ای از روش‌های قابل اعمال بر پایگاه داده‌های بزرگ و پیچیده به منظور کشف الگوهای پنهان و جالب توجه نهفته در میان داده‌ها، داده‌کاوی گفته می‌شود.

سرفصل های دوره داده کاوی با پایتون به شرح ذیل می باشد:

نصب و پیاده‌سازی محیط‌های عملیاتی

- آشنایی با ورژن‌های مختلف پایتون و نصب پایتون در لینوکس یا ویندوز همراه با نصب پکیج آناکوندا و آشنایی با پکیج‌های مهم
- نصب و ایجاد محیط اولیه در Eclipse و اجرای یک برنامه پایتون
- نصب و ایجاد محیط اولیه در Visual Studio Code و ایجاد یک برنامه پایتون
- نصب و ایجاد محیط اولیه در Jupyter و ساخت یک دفترچه پایتون

مفاهیم پایه داده‌ها و ریاضی و آماری

- داده و درک مفهوم ویژگی (Feature)، بعد (Dimension) و ماتریس (Matrix) و درک مفهوم تنسور (Tensor) و کاربرد آن در داده‌کاوی
- آشنایی و کار با کتابخانه‌های Numpy و Scipy برای انجام عملیات آماری
- آنالیز مولفه اصلی (PCA) و TSNE و کاربرد آن در نمایش داده‌ها و کاهش ابعاد
- بارگزاری داده‌ها و تعامل با داده‌ها با استفاده از کتابخانه‌ی Pandas

نمایش داده‌ها

- آشنایی با نمودارهای مختلف (Pie، Histogram، Bar، Line، Flow و...) و کاربرد هر یک

از آن‌ها

- نحوه‌ی نمایش هیستوگرام و کاربرد آن با کتابخانه‌ی Matplotlib
- نمایش داده‌ها به صورت تعاملی در کتابخانه‌ی Boken

طبقه‌بندی و رگرسیون و الگوریتم‌های مختلف آن

- آشنایی با نمونه داده‌های طبقه‌بندی و کاربردهای آن
- بررسی مجموعه داده‌های iris (تشخیص گل‌های زنبق از روی ویژگی‌ها)، MNIST (تشخیص تصاویر دست‌نوشته)، Boston Housing (قیمت‌گذاری هوشمند خانه) به عنوان نمونه‌های ساده و کاربردی
- معرفی روش‌ها و مراجع جمع‌آوری داده‌ها و استفاده از آن:
 - مثال پیش‌بینی هوشمند هزینه و تخمین ارزش کالا
 - مثال پیش‌بینی وضعیت هوا و هواشناسی
 - مثال کنترل ترافیک هوشمند با استفاده از داده‌های شهری
 - مثال تحلیل احساسات و استقبال/عدم استقبال کاربران از محصول یک فروشگاه با استفاده از کامنت‌های کاربران
 - مثال پیش‌بینی و توصیه محصول مورد نیاز کاربر در فروشگاه اینترنتی
 - مثال پیش‌بینی خرید کاربر با توجه به رفتار او در فروشگاه اینترنتی
 - مثال تشخیص هوشمند حملات هکرها به سرور
 - مثال پیش‌بینی هوشمند جرائم شهری و پیش‌گیری از وقوع جرم
 - مثال پیش‌بینی مصرف سوخت اتومبیل
- آشنایی و پیاده‌سازی طبقه‌بندی با الگوریتم نزدیک‌ترین همسایه (KNN) در پایتون

- آشنایی و پیاده سازی طبقه بندی با الگوریتم ماشین بردار پشتیبان (SVM) و آشنایی با

انواع مختلف پیاده سازی و پارامترهای آن در پایتون

- آشنایی و پیاده سازی طبقه بندی با الگوریتم بیز ساده (Naive Bayes) و انواع مختلف آن
با توجه به توزیع آماری داده ها

- بررسی درخت های تصمیم (Decision Trees) و پیاده سازی آن ها در حل مسائل
طبقه بندی در پایتون

- آشنایی و پیاده سازی طبقه بندی با الگوریتم های ترکیبی (AdaBoost ، RandomForest) در
پایتون

- آشنایی با الگوریتم های XGBoost و CatBoost و کتابخانه های XGBoost و CatBoost

- آشنایی با معیارهای مختلف ارزیابی کیفیت طبقه بندی

- Accuracy

- Precision

- Recall

- F1

- ROI AUC

خوشه بندی و الگوریتم های مختلف آن

- آشنایی با نمونه داده های خوشه بندی و حل مسائل کاربردی آن

- کاربرد و آشنایی با روش های عملی خوشه بندی

- مثال گروه بندی مشتریان (وبسایت و فروشگاه) با روش RFM و RFM مبتنی بر

زمان

- مثال گروه‌بندی تصاویر دست نوشته
- مثال گروه‌بندی هوشمند مطالب وبسایت بدون استفاده از ناظر
- مثال گروه‌بندی حملات هکرها به یک سرور
- آشنایی و پیاده‌سازی خوشه‌بندی با الگوریتم KMeans
- بررسی و پیاده‌سازی خوشه‌بندی با DBSCAN و OPTICS
- آشنایی با پیاده‌سازی DBSCAN سلسله مراتبی و کتابخانه HDBSCAN
- آشنایی و پیاده‌سازی خوشه‌بندی با الگوریتم MeanShift
- آشنایی و پیاده‌سازی خوشه‌بندی با الگوریتم سلسله مراتبی (Hierarchical Clustering)
- آشنایی و پیاده‌سازی خوشه‌بندی با الگوریتم طیفی (Spectral Clustering)
- آشنایی با روش‌های ارزیابی کیفیت خوشه‌ها

یہ پہلے باکاتر از تخصص ...