



جهاد دانشگاهی صنعتی شریف  
مرکز آموزش های تخصصی کاربردی

سرفصل های  
دوره آموزشی

داده کاوی با پایتون





جهاد دانشگاهی صنعتی شریف  
مرکز آموزش های تخصصی کاربردی

#### ▪ مقدماتی درباره داده کاوی با پایتون:

- طراحی مسائل و مجموعه داده هایی جهت شناخت داده، علم داده و کاربردهای آن در صنایع بیمه، بانک، بورس، طراحی نرم افزار با استفاده از علم داده، دیجیتال مارکتینگ هوشمند با داده کاوی، سیستم پیشنهاد دهنده وب سایت، متن کاوی در وب سایت و کاربردهای آن، تصویرکاوی و سیستم پیشنهاد موسیقی.
- بررسی ویژگی های زبان پایتون و پاسخ به این سوال که چرا از پایتون استفاده میکنیم؟ بررسی امکانات و زیرساخت های زبان پایتون به همراه جزئیات پیاده سازی برخی از قسمت ها با زبان (C)
- آشنایی عمومی با کتابخانه های موجود در زبان پایتون جهت انجام عملیات داده کاوی (Scikit Learn، Tensorflow، Py Torch، Numpy، Pandas، Matplotlib و ...)
- آشنایی کلی با حوزه ی کلان داده (Big Data)، هوش نرم و چهارچوب های مورد استفاده آن به همراه کاربرد ارتباط با علم داده

#### ▪ نصب و پیاده سازی محیط های عملیاتی:

- آشنایی با ورژن های مختلف پایتون و نصب پایتون در لینوکس یا ویندوز همراه با نصب پکیج آناکوندا و آشنایی با پکیج های مهم
- 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 داده ها و ریاضی و آماری:

- داده و درک مفهوم ویژگی (Feature)، بعد (Dimension) و ماتریس (Matrix) و درک مفهوم تانسور (Tensor) و کاربرد آن در داده کاوی



جهاد دانشگاهی صنعتی شریف  
مرکز آموزش های تخصصی کاربردی

- آشنایی و کار با کتابخانه ی Numpy و Scipy برای انجام عملیات آماری
- آنالیز مولفه اصلی (PCA) و TSNE و کاربرد آن در نمایش داده ها و کاهش ابعاد
- بارگزاری داده ها و تعامل با داده ها با استفاده از کتابخانه ی Pandas

#### ■ نمایش داده ها:

- آشنایی با نمودارهای مختلف (Pie، Histogram، Bar، Line، Flow و...) و کاربرد هر یک از آنها
- نحوه ی نمایش هیستوگرام و کاربرد آن با کتابخانه ی Matplotlib
- نمایش داده ها به صورت تعاملی در کتابخانه ی Boken

#### ■ طبقه بندی و رگرسیون و الگوریتم های مختلف آن:

- آشنایی با نمونه داده های طبقه بندی و کاربردهای آن
- بررسی مجموعه داده های iris (تشخیص گل های زنبق از روی ویژگی ها)، MNIST (تشخیص تصاویر دست نوشته)، Boston Housing (قیمت گذاری هوشمند خانه) به عنوان نمونه های ساده و کاربردی
- معرفی روش ها و مراجع جمع آوری داده ها و استفاده از آن
  - مثال پیش بینی هوشمند هزینه و تخمین ارزش کالا
  - مثال پیش بینی وضعیت هوا و هواشناسی
  - مثال کنترل ترافیک هوشمند با استفاده از داده های شهری
  - مثال تحلیل احساسات و استقبال/عدم استقبال کاربران از محصول یک فروشگاه با استفاده از کامنت های کاربران
- مثال پیش بینی و توصیه محصول مورد نیاز کاربر در فروشگاه اینترنتی
- مثال پیش بینی خرید کاربر با توجه به رفتار او در فروشگاه اینترنتی



جهاد دانشگاهی صنعتی شریف  
مرکز آموزش های تخصصی کاربردی

- مثال تشخیص هوشمند حملات هکرها به سرور
- مثال پیشبینی هوشمند جرائم شهری و پیشگیری از وقوع جرم
- مثال پیش بینی مصرف سوخت اتومبیل
  
- آشنایی و پیاده سازی طبقه بندی 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
- بررسی درخت های تصمیم (Decision Trees) و پیاده سازی آنها در حل مسائل طبقه بندی در پایتون
  
- آشنایی و پیاده سازی طبقه بندی با الگوریتم های ترکیبی (RandomForest، AdaBoost و...) 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
  - و ...

#### ▪ 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 (وب سایت و فروشگاه) با روش RFM و RFM مبتنی بر زمان



جهاد دانشگاهی صنعتی شریف  
مرکز آموزش های تخصصی کاربردی

- مثال گروه بندی تصاویر دست نوشته
- مثال گروه بندی هوشمند مطالب وب سایت بدون استفاده از ناظر
- مثال گروه بندی حملات هکرها به یک سرور
- آشنایی و پیاده سازی خوشه بندی با الگوریتم KMeans
- بررسی و پیاده سازی خوشه بندی با DBSCAN
- آشنایی با پیاده سازی DBSCAN سلسله مراتبی و کتابخانه ی HDBSCAN
- آشنایی و پیاده سازی خوشه بندی با الگوریتم MeanShift
- آشنایی و پیاده سازی خوشه بندی با الگوریتم سلسله مراتبی (Hierarchical Clustering)
- آشنایی و پیاده سازی خوشه بندی با الگوریتم طیفی (Spectral Clustering)
- آشنایی با روش های ارزیابی کیفیت خوشه ها
- Silhouette، کالینسکلی و ...

▪ متوازن سازی داده ها:

- الگوریتم های OverSampling ، SMOTE ، UnderSampling و ...
- آموزش کار با گوگل Colab و اجرای برنامه ها بر روی سرورهای Googlee
- آموزش کار با وب سایت Kaggle و کسب تجربه و رزومه
- کاهش ابعاد داده ها و الگوریتم های آن:  
- PCA ، UMAP، TSNE، KernelPCA



جهاد دانشگاهی صنعتی شریف  
مرکز آموزش های تخصصی کاربردی

تصویرکاوی و استفاده از تکنیک های پردازش تصویر دیجیتال (Histogram، HueMoments) و

(Haralick) در طبقه بندی و داده کاوی تصاویر