



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

سرفصل های
دوره آموزشی

دانشمند داده (Data Scientist)





جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

ریز سرفصل های دوره آموزشی دانشمند داده به شرح ذیل می باشد:

تحلیل اکتشافی

بخش اول مفاهیم ایده ها و ساختار:

- خاستگاه و اهمیت تحلیل اکتشافی
- مستندسازی تحلیل
- ساختار تحلیل داده
- تحلیل داده نظام مند

بخش دوم *Data Wrangling*:

- *Discovering*
- *Structuring*
- *Cleaning*
- *Enriching*
- *Validating*
- *Publishing*

تحلیل توصیفی داده ها بر اساس نوع داده و شناسایی ارتباطات دو و چند متغیره:

- تحلیل داده اسمی
- تحلیل داده کمی
- خلاصه سازی داده ها
- شناسایی ارتباطات دو و چند متغیره
- *Binning* و ساخت متغیر جدید



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

▪ بخش سوم آزمون فرض و *A/B test* :

- تعریف فرض صفر
- تعریف فرض جایگزین
- سطح معنی داری
- مقدار بحرانی
- تصمیم گیری و تفسیر نتیجه

▪ بخش چهارم ارائه نتایج:

- *Data Story Telling*
- *Resonate*

▪ بخش پنجم تحلیل داده بر مبنای شواهد (*Evidence based*) و اعتبارسنجی تحلیل داده:

- مخاطرات *EDA* و راهکارهای پاسخگویی به آنها
- تکرار پذیری
- تعمیم پذیری

▪ بخش ششم پروژه نمونه



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

داده کاوی با پایتون

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، تصویر کاوی 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



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

▪ 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 تنسور (**Tensor**) و کاربرد آن در داده کاوی
- 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 (**Flow**، **Line**، **Bar**، **Histogram**، **Pie** 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**



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

▪ 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
 - مثال تشخیص هوشمند حملات هکرها به سرور
 - مثال پیش بینی 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 (*AdaBoost*، *RandomForest*) و... در پایتون مرکز آموزش های تخصصی کاربردی
- 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
 - 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 *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 خوشه بندی طیفی (*Spectral Clustering*)
- آشنایی با روش های ارزیابی کیفیت خوشه ها
 - *Silhouette*، کالینسکی و ...



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

- متوازن سازی داده ها:
 - الگوریتم های *UnderSampling* ، *SMOTE* ، *OverSampling* و
- آموزش کار با گوگل *Colab* و اجرای برنامه ها بر روی سرورهای *Google*
- آموزش کار با وب سایت *Kaggle* و کسب تجربه و رزومه
- کاهش ابعاد داده ها و الگوریتم های آن:
 - *KernelPCA* ، *TSNE*، *UMAP* ، *PCA*
- تصویر کاوی و استفاده از تکنیک های پردازش تصویر دیجیتال (*HueMoments*، *Histogram* و *Haralick*) در طبقه بندی و داده کاوی تصاویر



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

Deep Learning

■ قسمت اول: مقدمات

- پایتون و کتابخانه های علمی مورد نیاز
- *OpenCV* در پایتون
- یادگیری ماشین چیست
- طبقه بندی و رگرسیون
- شناخت تصویر
- پیاده سازی طبقه بندی در پایتون

■ قسمت دوم: شبکه های عصبی

- نوران و پرسپترون
- تابع فعالیت
- یادگیری و بهینه سازها
- پیاده سازی شبکه های عصبی در *Keras*
- طبقه بندی اعداد دست نویس فارسی
- *Dropout*



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

▪ قسمت سوم: شبکه های کانولوشنالی

- شبکه های کانولوشنالی - مباحث تئوری
- شبکه های کانولوشنالی - عملی
- شبکه عصبی های کانولوشنال عمیق حائز رتبه و معروف
- انتقال یادگیری
- رگرسیون
- مدل های چند ورودی/چند خروجی در کراس *functional API* در کراس

▪ قسمت چهارم: مباحث تکمیلی یادگیری

- *Batch-Norm*
- *Learning-Rate Decay*
- *Tensorboard*
- *keras callbacks*
- *imbalanced data*

▪ قسمت پنجم: سیستم های پیشنهاد دهنده

- پیاده سازی *Book Recommendation System*

▪ قسمت ششم: شبکه های خودرمزگذار (*Auto-encoder*)

- *Up-sampling with Transposed Convolution*
- *Autoencoders*



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

▪ قسمت هفتم: **GAN**

- **Gans**
- **Conditional Gans**
- **Pix2Pix**
- **Cycle Gan**
- و ...

▪ قسمت هشتم: شبکه های بازگشتی

- شبکه های **LSTM ، GRU ، RNN**
- طبقه بندی متون
- طبقه بندی ویدیو
- **CTC loss** و مثال **OCR** بدون **segmentation**

▪ قسمت نهم: مباحث ویژه با توجه به علاقه و سرعت کلاس

- **Style transfer**
- **Deep dream**
- رنگی کردن تصویر
- بازشناسی چهره
- طبقه بندی - محلی سازی و شناسایی نظیر **SSD** و **Yolo** و ...