



سرفصل های دوره

متخصص علم داده با پایتون

Python Data Science

مدرس :

مسعود کاویانی

<https://www.7learn.com/experts/python-data-science>

گام اول : مقدمات

شناخت داده، علم داده و کاربردهای آن در صنعت

- صنعت بیمه، بانک، بورس
- طراحی نرم افزار با استفاده از علم داده
- دیجیتال مارکتینگ هوشمند با داده کاوی
- سیستم پیشنهاددهنده وبسایت
- متن کاوی در وبسایت و کاربردهای آن
- تصویرکاوی
- سیستم پیشنهاد موسیقی
-

آشنایی با علم هوش مصنوعی و قسمت های مختلف این علم

- الگوریتم های یادگیری ماشین و داده کاوی
- الگوریتم های بهینه سازی
- الگوریتم های هوشمند در کلان داده
- بررسی زبان های برنامه نویسی مرتبط با علم داده و بازار کار آنها
- پایتون
- R
- جاوا
- Scala

JavaScript •

C++ •

چرا از پایتون استفاده می‌کنیم؟

بررسی امکانات و زیرساخت‌های زبان پایتون به همراه جزئیات پیاده‌سازی برخی از قسمت‌ها با زبان C) آشنایی عمومی با کتابخانه‌های موجود در زبان پایتون جهت انجام عملیات داده‌کاوی (Scikit Learn، Tensorflow، PyTorch، Numpy، Pandas، Matplotlib و...) مفهوم کلی هوش تجاری و کاربرد آن در اجرا و پیاده‌سازی یک داشبورد مدیریتی در صنعت بانک‌داری آشنایی کلی با حوزه‌ی کلان‌داده (Big Data) و چهارچوب‌های مورد استفاده آن به همراه کاربرد ارتباط با علم‌داده تکنولوژی‌ی داکر جهت پیاده‌سازی پروژه‌های علم‌داده

گام دوم : نصب و پیاده سازی محیط‌های عملیاتی

آشنایی با ورژن‌های مختلف پایتون
نصب پایتون در لینوکس یا ویندوز همراه با نصب پکیج آناکوندا و آشنایی
با پکیج‌های مهم
نصب کتابخانه‌های یادگیری عمیق و یادگیری تقویتی

- Tensorflow
- Tensorflow-GPU
- Keras
- Keras-rl
- OpenAI GYM

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
نصب محیط عملیاتی داده‌کاوی بر بستر داکر (Docker) و اجرای آن

گام سوم : مفاهیم پایه داده ها و ریاضی و آماری

داده و درک مفهوم ویژگی (Feature)، بعد (Dimension) و ماتریس (Matrix)

درک مفهوم تانسور (Tensor) و کاربرد آن در داده‌کاوی

- پیاده‌سازی نمونه کاربردی تبدیل تصاویر تشخیص پلاک اتومبیل و تشخیص یک عنصر (مثلا یک انسان) در یک تصویر مناسب برای پیاده‌سازی ماشین‌های خودران)

آشنایی و کار با کتابخانه‌ی Numpy و Scipy برای انجام عملیات آماری

- حل برخی مسائل بهینه‌سازی با ریاضی در صنعت با Numpy

مقدمات جبر خطی (Linear Algebra)

- پیاده‌سازی عملیات پایه‌ی جبر خطی در پایتون

آنالیز مولفه اصلی (PCA) و TSNE و کاربرد آن در نمایش داده‌ها و کاهش ابعاد

- مثال کاربردی کاهش ابعاد در ساخت داشبورد مدیریتی

بارگزاری داده‌ها و تعامل با داده‌ها با استفاده از کتابخانه‌ی Pandas

- بارگزاری فرمت‌های مختلف داده‌ها از طریق یک یا چند فایل داده
- شناخت فرمت‌های مختلف داده مانند CSV، Excel، TSV، Txt و ارتباط و بارگزاری داده‌ها در پایتون با توجه به کتابخانه‌های رابطه‌ای مانند MySQL، MongoDB و...

- مثال خواندن داده‌های یک وب‌سایت فروشگاهی و وردپرس

گام چهارم : نمایش داده‌ها

عناوین این فصل، در راستای پیاده‌سازی یک داشبورد مدیریتی تعاملی، بر اساس KPI های یک سازمان انجام می‌شود.

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

نحوی نمایش هیستوگرام و کاربرد آن با کتابخانه Matplotlib و Seaborn

- پیاده‌سازی مدل کاربران در وبسایت

نحوی نمایش نمودار نقطه‌ای و کاربرد آن در کتابخانه Matplotlib و Seaborn

- پیاده‌سازی نقطه‌ی ورود کاربران به یک وبسایت

نحوی نمایش نمودار دونات و پای و کاربرد آن در کتابخانه Matplotlib و Seaborn

- پیاده‌سازی مدل اطلاعات مربوط به کاربران یک وبسایت

آشنایی با نمایش داده‌ها به صورت سه بعدی در Axe3D

نمایش داده‌ها به صورت تعاملی در کتابخانه Boken

- اجرای یک پروژه داشبورد تعاملی ساده برای داده‌های یک شرکت

گام پنجم : تست‌های آماری و تست‌های فرضیه

عناوین این فصل در راستای پیاده‌سازی کاربردی فرضیه‌های آماری در مورد داده‌هایی است که در کسب‌وکارها مورد استفاده قرار می‌گیرند. کتابخانه‌های `Numpy`، `Pandas`، `StatsModel` و `Prophet` کتابخانه‌های اصل هستند که در این فصل مورد استفاده قرار می‌گیرند. البته از کتابخانه‌های دیگر هم حسب نیاز استفاده می‌شود.

روش انجام تست‌های آماری جهت تحلیل داده

- مثال تحلیل رفتار کاربران یک اپلیکیشن
- تست فرضیه صفر و تست فرضیه یک
- مثال بهبود طراحی کمپین دیجیتال مارکتینگ
- روش انجام A/B تست در یک وبسایت فروشگاهی و تحلیل صفحه‌ی فرود (landing page)
- توزیع‌های مختلف آماری و آشنایی با کاربرد آن‌ها در تحلیل داده‌ها (گوسی، برنولی، طبقه‌ای، دریکله، توزیع F، توزیع T و...)
- مثال رفتار مشتریان اپلیکیشن
- انجام فرضیه‌ها با توجه به توزیع‌های آماری و کاربرد آن در داده‌کاوی
- بررسی P-Value و استفاده از آن در تست‌های آماری داده‌کاوی
- مثال بررسی رفتار کاربران در اپلیکیشن موبایل

بررسی T-Test و انجام انواع تست‌های T

• تست‌های One Sample t-test، Two Sample t-Test

بررسی تست‌های آماری مستقل از توزیع مانند Wilcoxon

بررسی تست‌های مکمل و نمونه کاربردی آن‌ها مانند Shapiro Test،

Fisher's F-Test، Chi-squared

• کاربرد تست‌های مکمل در تحلیل داده و کاهش ابعاد داده در طراحی

داشبورد مدیریتی

آنالیز سری‌های زمانی با کتابخانه‌ی Pandas و آنالیز سری‌های ARIMA و

SARIMA

• مثال پیش‌بینی داده‌های خریداران و مشتریان

• مثال پیش‌بینی زمان شلوغی خط مترو

آنالیز سری‌های زمانی با کتابخانه‌ی Prophet و مدل زنجیره‌ی مارکوف

• مثال تحلیل و پیش‌بینی آلودگی هوا

گام ششم : انبار داده‌ها و ساخت مکعب داده

عناوین این فصل در راستای ایجاد یک انبار داده جهت راه‌اندازی یک داشبورد مدیریتی در یک فروشگاه زنجیره‌ای است.

تفاوت انواع پردازش برخط (تراکنشی - OLTP و تحلیلی - OLAP) انواع مکعب داده بر روی داده‌های OLAP (مانند Snowflake، Start و سورفلکی)

- ساخت یک مکعب داده و ایجاد یک داشبورد مدیریتی تعاملی ساده

در پایتون با Boken و Pandas

اعمال Drill Down و Roll Up و پیاده‌سازی آن‌ها در پایتون

نحوی ذخیره‌سازی داده‌ها به صورت جریان داده (Stream Processing)

گام هفتم : طبقه‌بندی و رگرسیون و الگوریتم‌های مختلف آن

عناوین این فصل در راستای تسلط بر عملیات داده‌کاوی به صورت نظارت شده است. در همین راستا، نمونه‌های کاربردی به همراه کتابخانه‌های مختلف فراگرفته می‌شود. کتابخانه‌ی اصلی در این فصل Scikit Learn در پایتون است. هر چند کتابخانه‌های مختلف دیگر نیز حسب نیاز آموزش داده خواهند شد.

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 پیاده‌سازی طبقه‌بندی با الگوریتم ماشین بردار پشتیبان (SVM) و آشنایی با انواع مختلف پیاده‌سازی و پارامترهای آن در پایتون
- آشنایی و پیاده‌سازی طبقه‌بندی با الگوریتم بیز ساده (Naive Bayes) و انواع مختلف آن با توجه به توزیع آماری داده‌ها
- بررسی درخت‌های تصمیم (Decision Trees) و پیاده‌سازی آن‌ها در حل مسائل طبقه‌بندی در پایتون
- آشنایی و پیاده‌سازی طبقه‌بندی با الگوریتم‌های ترکیبی (AdaBoost، RandomForest و...) در پایتون
- آشنایی با الگوریتم‌های XGBoost و CatBoost و کتابخانه‌های CatBoost و XGBoost
- آشنایی با معیارهای مختلف ارزیابی کیفیت طبقه‌بندی

- Accuracy
- Precision
- Recall
- F1

گام هشتم : خوشه‌بندی و الگوریتم‌های مختلف آن

عناوین این فصل در راستای تسلط بر یادگیری غیرنظارت شده در داده‌کاوی است. در این فصل از کتابخانه‌ی `scikit learn` و کتابخانه‌های دیگر جهت انجام عملیات خوشه‌بندی استفاده خواهیم کرد.

- آشنایی با نمونه داده‌های خوشه‌بندی و حل مسائل کاربردی آن
- کاربرد و آشنایی با روش‌های عملی خوشه‌بندی
- مثال گروه‌بندی 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 کتابخانه‌ی HDBSCAN
- آشنایی و پیاده‌سازی خوشه‌بندی با الگوریتم MeanShift
- آشنایی و پیاده‌سازی خوشه‌بندی با الگوریتم سلسله مراتبی (Hierarchical Clustering)

آشنایی و پیاده‌سازی خوشه‌بندی با الگوریتم خوشه‌بندی طیفی (Spectral Clustering)

آشنایی با روش‌های ارزیابی کیفیت خوشه‌ها

• Silhouette

• کالینسکلی

• و...

گام نهم : شبکه‌های عصبی و شبکه‌های عصبی عمیق

عناوین این بخش جهت تسلط به شبکه‌های عصبی جهت پیاده‌سازی در حوزه یادگیری نظارت شده، غیرنظارت شده و نیمه نظارت شده مورد استفاده قرار می‌گیرد. در این فصل بر روی پیاده‌سازی عناصر مختلف برای ساخت ماشین‌های خودران کار می‌کنیم. کتابخانه‌های Tensorflow ، PyTorch و Keras از کتابخانه‌هایی هستند که در این فصل با آن‌ها کار خواهیم کرد.

معماری شبکه‌های عصبی و پیاده‌سازی یک شبکه ساده در پایتون
آشنایی با معماری‌های مختلف و لایه‌های متفاوت در شبکه‌های عصبی
معماری پس‌انتشار خطا (Back Propagation of Error) و کاربرد آن در
شبکه عصبی

آشنایی با توابع فعال‌سازی مختلف و کاربرد آن‌ها

- Sigmoid
- Relu
- Tanh
- ...

آشنایی با انواع معماری شبکه عصبی عمیق
معماری شبکه‌ی خود رمز نگار (Auto Encoder) و شبکه‌های عصبی مولد
(GAN) و پیاده‌سازی

- مثال ساخت تشخیص دهنده در ماشین‌های خودران معماری شبکه‌های عصبی پیچیده (CNN) و پیاده‌سازی
- مثال تشخیص عناصر در تصویر ضبط شده در یک ماشین خودران معماری شبکه‌های عصبی ترتیبی و RNN و LSTM جهت تحلیل داده‌های زمانی و پیاده‌سازی
- مثال عملی ترجمه‌ی متن به صورت هوشمند

گام دهم : بهینه‌سازی و الگوریتم‌های هوشمند بهینه‌ساز

عناوین این فصل جهت ساخت مدل‌هایی انجام می‌شوند که می‌توانند مسائل مختلف در صنعت را بهینه‌سازی نمایند. در این فصل بر یک سیستم برنامه‌ریز هوشمند و همچنین سیستم انبارداری هوشمند با الگوریتم‌های فراابتکاری در محیط پایتون کار می‌شود.

آشنایی با مفهوم بهینه‌سازی و حل مسائل بهینه‌سازی پیچیده
مفهوم تابع برازش (Fitness Function) و کاربرد آن

- مسئله‌ی بهینه‌سازی فروش
 - مسئله‌ی بهینه‌سازی صفحه‌ی وب‌سایت/اپلیکیشن به صورت هوشمند و خودکار
 - مسئله‌ی بهینه‌سازی انبار و کالاهای موجود
- الگوریتم جستجوی تصادفی (Random Search) و تپه‌نوردی (Hill Climbing)
- الگوریتم بهینه‌سازی کلونی مورچگان (ACO)
- الگوریتم بهینه‌سازی جستجوی ممنوعه (Tabu Search)
- الگوریتم بهینه‌سازی سگ‌های وحشی آفریقا
- الگوریتم بهینه‌سازی مگس میوه
- الگوریتم بهینه‌سازی زنبورهای هوشمند (ABC)

گام یازدهم : شبکه‌های عصبی عمیق تقویتی (Reinforcement Deep Neural Networks)

عناوین این فصل جهت پیاده‌سازی مفاهیم و تسلط بر شبکه‌های عصبی عمیق تقویتی، تشکیل شده است. در این فصل بر روی مثال‌هایی از طراحی یک عامل یادگیرنده جهت ماشین‌های خودران، انجام یک بازی توسط شبکه و یادگیری در محیط تقویتی، کار می‌شود. در این فصل از چهارچوب‌های Keras-RL و OpenAL GYM استفاده می‌شود و کتابخانه‌های دیگر نیز در مورد نیاز، مورد استفاده قرار می‌گیرند.

آشنایی با مفاهیم یادگیری تقویتی و عامل یادگیرنده

آشنایی با شبکه‌ی عصبی عمیق Deep Q Network

آشنایی با شبکه‌های مبتنی بر Policy در یادگیری تقویتی

- ساخت عوامل ماشین‌خودران
- ساخت یک agent جهت یادگیری تعادل یک آونگ
- ساخت یک agent جهت یادگیری یک بازی کامپیوتری ساده