



سازمان جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی



دوره جامع علم داده (Data Science)



www.sharif.ac



۰۲۱-۶۷۶۴۱۹۹۹

جهان امروز، جهان داده است و ذخیره سازی و استفاده بهینه از داده ها یکی از چالش های اصلی سازمان ها در یک دهه گذشته بوده است. از این رو در سال های اخیر، تمرکز اصلی متخصصان داده بر روی ساختن چهارچوب ها و راه حل هایی برای مدیریت داده ها بوده است. آموزش علم داده (Data Science) در حقیقت حلقه اتصال و هدایت گر این وضعیت است. در ادامه قرار است با آموزش علم داده بیشتر آشنا شویم.

سرفصل های دوره جامع علم داده به شرح ذیل می باشد:

SQL Server

- آشنایی با دوره آموزشی و ابزارهای ارتباطی و اشتراک فایل مورد نیاز
- آشنایی عمومی با مفهوم دیتا و دیتابیس
- معرفی انواع دیتابیس ها
- آشنایی با مدل های توزیع دیتا
- نصب و راه اندازی SQL Server و راه اندازی دیتابیس تمرینی
- آشنایی با محیط SSMS و دیتابیس تمرینی و ابزار های مورد نیاز
- آموزش دستور Select
- آموزش نمایش داده ها با TOP
- آموزش دستور Concat و ترکیب ستون ها
- آموزش استفاده از انواع نام مستعار در کوئری نویسی
- آموزش نمایش داده های غیر تکراری (Distinct)
- آموزش نوشتن شرط و فیلتر در Clause Where
- آشنایی با انواع Data Type در SQL Server
- آموزش استفاده از دستور Between
- آموزش استفاده از عملگر شباهت متنی Like
- آشنایی با مفهوم NULL و فیلتر کردن آن
- مرتب سازی داده ها با Order By
- آشنایی با مفهوم Set Base

- آموزش استفاده از Except و Intersect
- استفاده از Count برای شمارش
- ایجاد جدول برای ذخیره سازی داده
- آشنایی با مفهوم Incremental و ایجاد ترتیب اتوماتیک دادهای ورودی با Identity
- انتقال داده به جداول
- آشنایی با انواع کلید و روابط بین جداول و ایجاد آنها
- کشف و مشاهده روابط داده ها و جداول با Diagram و نحوه ایجاد آن.
- آشنایی و استفاده از دستور و مفهوم Union و Union All
- تغییر و به روز رسانی داده ها با Update
- پاک کردن داده ها با Delete یا Truncate و بررسی تفاوت ها
- استفاده از مفهوم مجموعه ها، اشتراک و اجتماع در شرط نویسی با AND , OR
- استفاده از توابع شمارش و محاسبه در Server SQL با AVG,Max,Min,Count
- آشنایی با WildCard
- در ادامه مباحث مجموعه ها، بررسی وجود و عدم وجود با استفاده از IN و Not in
- تعریف Normalization داده ها و کاربرد های آن
- ادغام داده های جداول با مفهوم JOIN و بررسی دیاگرام های مرتبط:
- Inner JOIN
- Left JOIN
- Right JOIN
- Full Outer JOIN
- آشنایی و استفاده از توابع تغییر Data Type شامل Convert , Cast
- آشنایی با موضوع تاریخ و استفاده از توابع تاریخی و مفهوم زمان:
- Getdate
- DateAdd

Format

DateDiff

DateName

Day, Month, Year

▪ آشنایی با توابع متنی و استفاده از آنها:

Left

Right

Replace

Trim

Space

Stuff

Len

PATINDEX

CHARINDEX

SUBSTRING

▪ آشنایی با توابع ریاضی و استفاده از آنها:

CEILING

Floor

EXP

Log

Power

Rand

Round

SQRT

▪ بررسی انواع حالت انتقال داده ها به جداول و تفاوت ها

Insert Into

Select Into

- آشنایی و استفاده از SubQuery ها و نوشتن کوئری های تحلیلی:

Self-Contained Subqueries

Nested Subquery

Correlated Subqueries

Drived Table

- آشنایی و استفاده از انواع جداول موقت و کاربردهای فنی آن:

Local Temp Table

Global Temp Table

- آشنایی مقدماتی با Window Function

- آشنایی با توابع دسته بندی و تحلیل داده ها:

Row_Number

Lag

Lead

First_value

Last_value

Rank

Dens_rank

- آشنایی با نحوه ایجاد و استفاده از نمایه ها View

- آشنایی و استفاده از مفهوم Pivot در تحلیل داده ها

- آشنایی مقدماتی با مباحث index و performance

تحلیل داده با JMP SAS و تحلیل اکتشافی

- Descriptive Data Analysis
 - Types of Variables and Data
 - Measures of Central Tendency
 - Mean, Median, Mode
 - Measures of Dispersion (Variability)
 - Range, Interquartile Range (IQR), Variance, Standard Deviation, Coefficient of Variation
 - Measures of Shape
 - Skewness, Kurtosis
 - Graphical Data Representation Methods: Histogram, Box Plot, Bar Chart
- An Overview of the Concept of Probability
 - Conditional Probability, Independent Events, Mutually Exclusive Events, Law of Total Probability, and Bayes' Theorem
- Random Variable and Probability Functions
 - Probability Distribution and Probability Density
 - Introduction to Some Common Distributions: Uniform, Bernoulli, Binomial, Poisson, Exponential, and Normal
 - Normality of Data for a Random Variable
 - Transforming Non-Normal Data to Normal Data Using Transformations (Box-Cox, Johnson)
- Multivariate Distributions
 - Examining the Relationship Between Two Random Variables
 - The Concept of Correlation and Interpretation of the Correlation Coefficient

Tableau

- آشنایی با نرم افزار Getting Started
- آماده سازی داده ها Connecting to Data
- بصری سازی و ویژوالیزشن Visual Analytics
- معرفی ویژوال های کاربردی
- ساخت داشبوردهای مدیریتی Dashboards and Stories
- اشتراک گذاری داشبوردها و گزارش ها روی موبایل و تبلت

Microsoft Power BI

- معرفی سرویس ها و ابزارهای Power BI و کارکرد آنها
- تشریح مراحل پنجگانه سیکل گزارش های تحلیلی و داشبوردها با ابزار Power BI
- مروری بر دریافت و اصلاح داده ها از منابع مختلف داده
- مروری بر مدل سازی داده ها و معرفی توابع DAX برای ساخت محاسبه های کاربردی
- مروری بر مصورسازی (تحلیل) داده ها و معرفی ویژوال های کاربردی
- معرفی Power BI Service برای به اشتراک گذاری داشبوردها
- اشتراک گذاری گزارش ها به کمک Power BI Report Server
- اشتراک گذاری داشبوردها و گزارش ها روی موبایل و تبلت

R For Data Engineering

Installation:

- Data Types
- Control Flow in R
- Vectorized Thinking
- introduction to packages
- Introduction to Date & Time manipulation
- Introduction to String Manipulation
- Importing Data
 - Importing Data from Flat Files
 - Importing Excel Files
- Data Manipulation
 - Data Transformation

Python For Data Engineering

Installation:

- Python & Anaconda installation
- Basic Data in python
- Variables in python
- package Management
- introduction to numpy
- introduction to pandas
- introduction to control flows
- data importing in python

- importing data from text files
- importing data from csv
- importing data from excel
- data manipulation
 - datetime manipulation fundamentals
 - string manipulation fundamentals
 - table manipulation fundamentals
- introduction to matplotlib

BIG DATA

- What is Big Data?
- How to gain insights from Big Data?
- Learn the important concepts, tools, standards, and uses of Big Data.
- Learn about Big Data implementation.
- Learn how to manage Big Data.
- Examine Big Data analytics.
- Explore the different elements of Big Data.
- Available solutions for Big Data
- Big Data Architectures
 - Data Lake
 - Lambda Architecture
 - Kappa Architecture
- Big Data Collection Solutions
 - Getting Data from RDBMS
 - Getting Streaming Data

- Gathering Logs Data
- Big Data Storage Solutions
 - Distributed File Systems
 - NoSQL Databases
 - Object Storage
- Big Data Processing Frameworks
 - Batch, streaming and real-time Processing Engines
 - NoSQL Query Engines
- Streaming Data
- SQL on Big Data

Python & R for Machine Learning

- مقدمه ای بر یادگیری ماشین
- الگوریتم KNN با python & R
- تحلیل رگرسیون logistic
- تحلیل Naive Bayes
- الگوریتم ماشین بردار پشتیبان
- الگوریتم درخت تصمیم
- روش های کشف قوانین انجمنی
- تحلیل رگرسیون خطی چند متغییره
- تحلیل رگرسیون غیر خطی (مدل سازی پیچیده)
- الگوریتم embedded learning
- الگوریتم قانون طبقه بندی
- الگوریتم های خوشه بندی

یک پل به بالاتر از تخصص ...

