

## معرفی دوره

# Fundamentals of Data Science

### 1- این دوره برای چه کسانی مناسب است؟

این دوره برای تمامی افرادی که علاقه‌مند به یادگیری تحلیل داده و هوش مصنوعی هستند مناسب است، لازم به ذکر است که با توجه به گسترده شدن علم داده و هوش مصنوعی اکثر رشته‌ها به این موضوع گره خورده‌اند و افرادی که قصد استفاده از قدرت هوش مصنوعی در فعالیت‌های شخصی و شغل خود را دارند می‌توانند از این فرصت استفاده کنند تا خلاقیت را وارد کسب و کار کنند.

همچنین این دوره برای دانشجویان تمامی رشته‌ها مانند ریاضی، علوم کامپیوتر، آمار، فیزیک، شیمی و... که قصد افزایش عمق شناخت از تحلیل داده را دارند مفید می‌باشد.

این دوره فقط برای افرادی بازخورد مناسب خواهد داشت که پیش زمینه مناسبی از برنامه نویسی به زبان پایتون و مفاهیم اون داشته باشند، ولی احتیاجی نیست پیش زمینه‌ای از آشنایی با ابزارها و مباحث مربوط به تحلیل داده داشته باشند.

### 2- هدف این دوره چیست؟

هدف اصلی این دوره مرور کامل و جامع تحلیل علم داده و شناخت استفاده‌های روزمره و واقعی این علم است، مباحث دوره بازه گسترده‌ای از تمامی مراحل کار با داده را شامل می‌شود، از جمع آوری، پیش پردازش تا تحلیل اکتشافی و تحلیل داده همچنین مباحث تکمیلی مانند دیتابیس، وب اسکریپینگ و مصور سازی داده را نیز شامل می‌شود.

این دوره طراحی شده است تا بنیان قوی و کاملی در زمینه‌ی تحلیل داده و بسیاری از مهارت‌های کاربردی مربوط به آن را برای دانش پژوهان و علاقه مندان به علم داده ایجاد کند به شکلی که در زمینه علمی و کسب و کار واقعی بتوانند به شکل موثری از آن استفاده کنند.

### 3- در این دوره چه چیزی کسب خواهید کرد؟

در انتهای این دوره شرکت‌کنندگان شناخت کافی نسبت به تمام مسیرهای پیش‌روی تحلیل علم داده پیدا خواهند کرد و می‌توانند از دانش خود در حل مسایل دنیای واقعی استفاده کنند. شرکت‌کنندگان تجربه کار عملی و تحلیل داده که شامل مدل‌های آماری، تحلیل همبستگی و مصور سازی داده می‌باشد را نیز فراخواهند گرفت. در نهایت شرکت‌کنندگان توانایی جمع‌آوری داده را به طور کامل برای استفاده در مدل‌های مختلف و پیش‌پردازش آن داده‌ها کسب خواهند کرد. با توجه به اینکه مسیر آموزش تحلیل داده به شکل کامل در این دوره تدریس می‌شود، بعد از اتمام این دوره علاقه‌مندان توانایی به کار بردن روش‌های تحلیلی در علم داده و پیاده‌سازی آن را در مسائل مدنظر خود خواهند داشت.

## مباحث دوره:

### فصل اول: داده‌ها

- داده چیست؟
- داده‌ها در مقابل اطلاعات
- انواع داده‌ها
  - داده‌های ساختاریافته (Structured data)
  - داده‌های جدولی (Tabular)
  - سری زمانی (Time Series)
  - گراف (Graph)
  - تصویر (Image)
  - صدا (Audio)
  - ویدئو (Video)
- داده‌های ساختارنیافته و نیمه ساختاریافته (Unstructured and Semi-structured data)
- داده‌های دیجیتال: قالب‌های ذخیره‌سازی (Digital data: storage formats)
  - متن (Text)
  - رمزگذاری‌ها (Encodings)
  - مجموعه حروف (Character sets)
  - CSV و قالب‌های جدولی دیگر
  - JOSN
  - XML
  - BMP و قالب‌های تصویر دیگر

WAV •

• پایگاه داده (Database)

• چرا به پایگاه داده نیاز داریم؟

• SQL

• پایگاه داده های NoSQL

• داده در کجا وجود دارد؟

## فصل دوم: جمع آوری داده ها (Data Collection)

• روش های جمع آوری داده ها

○ نظرسنجی ها (Surveys)

○ سیستم های بازخورد (Feedback Systems)

○ اسکرپینگ وب (Web scraping)

▪ معرفی اصطلاحات

▪ معرفی ساختار وب (Introduction to web)

▪ تکنولوژی های وب (Web Technologies)

▪ HTML و زبان های Markup دیگر

▪ تجزیه HTML (HTML parsing)

▪ HTTP

▪ API

▪ چالش های اسکرپینگ وب (Challenges of web scraping)

• وبسایت های Ajax (Ajax websites)

• Firewall

• محدودیت نرخ درخواست (Rate limits)

▪ چگونه با این چالش ها برخورد کنیم

▪ وب اسکرپینگ موازی و بهینه (Concurrent/Async Web scraping)

▪ داده های گراف و شبکه ها: جمع آوری وب بازگشتی (Graph data in web: Recursive web scraping)

Recursive web scraping)

## فصل 3: پیش پردازش و کاوش داده (Data Preprocessing and exploration)

• پیش پردازش (Preprocessing)

- یکپارچه سازی (Integration)
- مشخص کردن مساله (Problem identification)
- حذف کردن و پر کردن (Dropping and filling)
- تخمین زمانی سری زمانی (Time series interpolation)
- تمیز کردن متن (Text cleaning)
- پایپ لاین های پردازش داده (Transformation pipelines)
- کاوش و تجزیه و تحلیل داده (Exploration, and Data Analysis)
  - آمار معیارها (Statistical metrics)
  - تحلیل همبستگی (Correlation analysis)
  - آزمون فرضیه (Hypothesis testing)
  - بصری سازی (Visualization)
    - تحلیل تک متغیره (Univariate analysis)
    - تحلیل دو متغیره (Bivariate analysis)
    - کانال های ادراک (Perception channels)
    - تجزیه و تحلیل چند متغیره (Multivariate analysis)
    - تبدیل مثال های تک متغیره به چند متغیره (Converting bivariate examples to multivariate)
    - تکنیک های تجمیع (Aggregation techniques)
      - چند نوع نمودار شناخته شده
      - بصری سازی سه بعدی (3D visualization)
      - بصری سازی پویا (Realtime visualization)
- پیش پردازش برای مدل سازی (Preprocessing for modeling)
  - کدگذاری ویژگی ها
  - کاهش ابعاد
    - PCA
    - TSNE
  - هماهنگی و تغییر مقیاس داده (Normalization and scaling)
  - تطبیق نمونه های داده (افزایش نمونه، کاهش نمونه)
  - (Data resampling (Up-sampling, Down-sampling))

مباحثی که در هر جلسه درس داده خواهد شد:

این دوره شامل 15 جلسه 2 ساعته می باشد

Session	Topics
1	What is data? Data vs. information Types of data Storage formats Introduction to Database
2	Database cont. Where data resides? Introduction to data collection
3	Data collection methods Surveys Feedback systems Web Scraping, Introduction
4	Web Scraping cont.
5	Web Scraping cont.
6	Preprocessing Integration Problem identification Dropping and filling
7	Preprocessing cont. Time series processing Time series interpolation
8	Text cleaning and transformation Transformation pipelines
9	Exploration, and Data Analysis Statistical metrics Correlation analysis Hypothesis testing
10	Visualization Univariate analysis Bivariate analysis Perception channels Multivariate analysis Converting bivariate examples to multivariate
11	Visualization cont. Aggregation techniques Some well-known plot types
12	3D visualization Realtime visualization
13	Feature encoding Dimension reduction PCA TSNE
14	Normalization and scaling Data resampling (Up-sampling, Down-sampling)
15	Recap Real world miniproject