



سازمان جهاد دانشگاهی صنعتی شریف  
مرکز آموزش های تخصصی کاربردی



# متن کاوی با پایتون

امروزه علوم داده به یکی از جذابترین و پرترفدارین تخصص ها و مشاغل در ایران و دنیا تبدیل شده است. اما برای تبدیل شدن به یک متخصص علوم داده، تسلط بر مفاهیم و تکنیک های داده کاوی های معمول و شناخته شده یا همان کاوش دیتای ساختار یافته (Structured Data) کافی نیست و تسلط به تحلیل داده های بدون ساختار (Unstructured Data) امری ضروری محسوب می شود. پردازش و کاوش متون و زبان های طبیعی نیز یک جزء ازداده های ساختار نیافته می باشد. **متن کاوی** یا Text Mining به فرآیند داده کاوی Data Mining در متون و زبان های طبیعی گفته می شود.

سرفصل های دوره متن کاوی با پایتون به شرح ذیل می باشد:

- مقدمه ای بر متن کاوی
- آشنایی با کتابخانه nltk
- پیش پردازش متن
- مقدمه ای بر عبارات منظم
- ساخت بردار ویژگی ها با استفاده از کلمات
- رده بندی و خوشه بندی متن ها
- تحلیل احساسات
- متن کاوی فارسی
- تعبیه سازی کلمات و کاربرد یادگیری عمیق (deep learning) در متن کاوی

یہ پہلے باکاتر از تخصص ...