

مدرس: فریبا عزیزی	نام درس: داده کاوی	نیم سال: 991
رشته و مقطع تحصیلی: آمار-کارشناسی	پیش نیاز: احتمال ۲ و رگرسیون	نوع درس و تعداد واحد: اختیاری-۳ واحد
نحوه ارزیابی دانشجویان: کوئیز، آزمون های میان ترم، حل تمرین، پروژه R		تاریخ آزمون پایان ترم: ۹۹/۱۱/۷ تاریخ آزمون میان ترم: با هماهنگی دانشجویان
ایمیل استاد: fa.azizi@alzahra.ac.ir		
هدف کلی: آشنایی با مفاهیم اولیه داده کاوی، پالایش داده ها و الگوریتم های داده کاوی اهداف جزئی: آشنایی با روش های پیش پردازش داده ها، تحلیل اکتشافی داده ها، الگوریتم K-NN، درخت های تصمیم، خوشه بندی، شبکه های عصبی، قوانین انجمنی و ارزیابی مدل.		
مراجع: Larose, D. T., & Larose, C. D. (2014). Discovering knowledge in data: An Introduction to data mining (Vol.4.) John Wiley & Sons.		
	هفته	شرح درس
اول	معرفی درس، منابع، سرفصل ها، مقدمه ای بر داده کاوی و چکیده ای از کارهایی که می توان با داده کاوی انجام داد.	دوم
دوم	پیش پردازش داده ها، پاکسازی داده ها، نحوه برخورد با داده های گمشده و عجیب، تبدیل مقیاس متغیرها، دسته بندی متغیرهای عددی و باز طبقه بندی متغیرهای رسته ای	سوم
سوم	تحلیل اکتشافی داده ها، بررسی ارتباط بین متغیرها با ابزارهای گرافیکی و عددی و پیاده سازی آن بر روی مجموعه داده Churn	چهارم
چهارم	ادامه ای تحلیل اکتشافی داده ها، کشف زیرمجموعه ای مناسب از داده ها، ایجاد متغیرهای جدید کمی و کیفی، مروری بر روش های تحلیل تک متغیره، فاصله اطمینان، آزمون فرض میانگین و نسبت در یک جامعه	پنجم
پنجم	مروری بر روش های تحلیل چند متغیره، آزمون فرض برابری میانگین دو جامعه، برابری نسبت دو جامعه و همگنی واریانس ها مروری بر آنالیز واریانس و رگرسیون	ششم
ششم	معرفی روش های بانظر و بدون ناظر، مفاهیم اعتبارسنجی متقابل، بیش برآزش و کم برآزش و بایاس و واریانس	هفتم
هفتم	بررسی الگوریتم KNN و مسائل مربوط به ساخت مدل آن، تعریف تابع فاصله و تعریف انواع توابع ترکیب	هشتم
هشتم	بررسی الگوریتم درخت های تصمیم، الگوریتم CART و C4.5 و نحوه ای استخراج قواعد تصمیم	نهم
نهم	بررسی الگوریتم شبکه های عصبی، روش های کدگذاری ورودی ها، ویژگی های شبکه عصبی، معرفی لایه ها و گره ها و توابع فعال سازی	دهم
دهم	ادامه ای شبکه های عصبی، نحوه یادگیری شبکه عصبی، قوانین پس انتشار خطا، قوانین توقف الگوریتم و نرخ یادگیری الگوریتم	یازدهم
یازدهم	بررسی الگوریتم های خوشه بندی، الگوریتم سلسله مراتبی مبنی بر تفکیک و ترکیب، خوشه بندی غیر سلسله مراتبی k-means	دوازدهم
دوازدهم	قوانین انجمنی، کاوش قوانین انجمنی، روش های القای قانون تعمیم یافته، پیاده سازی الگوریتم	

سیزدهم	ادامه‌ی قوانین انجمنی، شروع مباحث مربوط به ارزیابی مدل، ارزیابی نتایج عمل توصیف
چهاردهم	ارزیابی برای عمل برآورد و پیش‌بینی، ارزیابی برای عمل طبقه‌بندی، تنظیم هزینه طبقه‌بندی غلط
پانزدهم	جلسات حل تمرین و رفع اشکال
شانزدهم	امتحانات میان‌ترم