



دانشکده بهداشت

دفتر توسعه آموزش

طرح دوره (Course Plan)

نام درس: اصول و روش‌های داده کاوی	کد درس: ۴۱۱۲۱۷
مقطع تحصیلی: کارشناسی ارشد	رشته تحصیلی: آمار زیستی
گروه آموزشی: آمار زیستی	پیش نیاز: محاسبات آماری با کامپیوتر
تعداد واحد: ۲ واحد نظری	محل برگزاری: دانشکده بهداشت
روز و ساعت برگزاری: دوشنبه ۱۰-۱۲	مسئول درس: دکتر سمانه صبوری
بازنگری طراحی اولیه	
هدف کلی دوره: استفاده از الگوریتم‌های آماری و یادگیری ماشین جهت کشف الگوهای رفتاری موجود در داده‌ها از اهداف این دوره می‌باشد که به وسیله آن درک بهتری از روابط موجود در داده‌ها حاصل می‌گردد. دانشجو در این درس تقابل بین مدل‌های آماری و مدل‌های مبتنی بر ماشین را خواهد آموخت. در انتهای دوره دانشجو باید بتواند با استفاده از نرم‌افزار پایتون و R تحلیل مناسی از الگوهای رفتاری پدیده‌ها ارائه دهد.	
اهداف اختصاصی دوره: ✓ آشنایی با مفهوم یادگیری با نظارت و بدون نظارت ✓ آشنایی با روش‌های انتخاب متغیر و Regularization ✓ آشنایی با روش‌های بر اساس درخت تصمیم ✓ آشنایی با الگوریتم‌های ماشین بردار پشتیبان ✓ آشنایی با الگوریتم‌های خوشه‌بندی ✓ آشنایی با روش‌های کاهش بعد در داده‌ها ✓ آشنایی با کاربرد نرم‌افزارهای R و پایتون در روش‌های یادگیری ماشین	
جدول ترتیب ارائه محتوای آموزشی در طول دوره	
جلسه	رئوس مطالب
اول	مفهوم داده‌کاوی، مفهوم داده با بعد بالا و مه داده
دوم	پیش پردازش، پاکسازی و آماده‌سازی داده‌ها
سوم	مفهوم یادگیری آماری (با نظارت/ بدون نظارت)
چهارم	آشنایی با روش‌های بصری سازی داده‌ها با کمک R و پایتون
پنجم	رگرسیون خطی، لجستیک و روش‌های کلاسیک انتخاب متغیر همچون گام به گام
ششم	آشنایی با الگوریتم‌های مورد استفاده در روش‌های رده بندی
هفتم	میان ترم
هشتم	روش‌های ارزیابی متقاطع (اعتبار درونی، اعتبار بیرونی و اعتبارسنجی متقابل) و ارزیابی عملکرد الگوریتم‌ها



دانشکده بهداشت

دفتر توسعه آموزش

طرح دوره (Course Plan)

(صحت، دقت، سطح زیر منحنی راک)	
روش‌های ریج، لاسو و الاستیک نت	نهم
درخت تصمیم	دهم
بگینگ، بوستینگ، جنگل تصادفی	یازدهم
کاربرد تحلیل مولفه اصلی و تحلیل عاملی در کاهش بعد	دوازدهم
کاربرد تحلیل مولفه اصلی و تحلیل عاملی در کاهش بعد: ادامه	سیزدهم
الگوریتم بردارهای پشتیبان خطی	چهاردهم
الگوریتم بردارهای پشتیبان غیر خطی و روش‌های Kernel	پانزدهم
شبکه عصبی مصنوعی	شانزدهم
شبکه عصبی مصنوعی: ادامه	هفدهم
روش‌های آموزشی (نحوه ارائه درس): آموزش از طریق سخنرانی و اسلاید، طرح پرسش در هر جلسه	
وسایل آموزشی: ویدئو پروژکتور، وایت بورد و مژیک	
وظایف و تکالیف دانشجوی: مرور مطالب بیان شده در جلسات قبلی، حل تمرینات تعیین شده، حضور فعال و مشارکت در فعالیت‌های کلاسی	
روش‌های ارزیابی دانشجوی (لطفا سهم هر روش بر حسب درصد نوشته شود): روش‌های ارزیابی در طول دوره: فعالیت‌های کلاسی (۲۰٪)، آزمون کتبی میان ترم (۲۰٪) روش‌های ارزیابی در انتهای دوره: آزمون کتبی پایان ترم (۶۰٪)	
منابع اصلی درس:	
۱ James G, Witten D, Hastie T, Tibshirani R, Taylor J. An introduction to statistical learning: With applications in python: Springer Nature; 2023.	
۲ Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer; 2017.	
۳ Lan H, Witten Eibe Frank . Mark A Hall. Data Mining : Practical Machine Learning Tools and Techniques, Elsevier. Last edition	