



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: چهارشنبه ۱۰-۱۲	مسئول درس: دکتر سمانه صوری
طراحی اولیه ☒ بازنگری ☐	
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 مناسبی 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، اهمیت، اهداف و کاربرد آن در داده‌های سلامت
دوم	پاکسازی و آماده‌سازی داده‌ها
سوم	مفهوم یادگیری آماری (با نظارت/ بدون نظارت)
چهارم	رگرسیون خطی، لجستیک و روش‌های کلاسیک انتخاب متغیر همچون گام به گام
پنجم	آشنایی با روش‌های رده بندی بر مبنای احتمال و غیر احتمالی
ششم	روش‌های ارزیابی متقاطع (اعتبار درونی، اعتبار بیرونی و اعتبارسنجی متقابل)
هفتم	ارزیابی عملکرد الگوریتم‌ها و برآوردها با معیارهای صحت، دقت و سطح منحنی راک
هشتم	روش‌های ریح، لاسو و الاستیک نت



دانشکده بهداشت

دفتر توسعه آموزش

طرح دوره (Course Plan)

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	روش‌های خوشه‌بندی
شانزدهم	روش‌های خوشه‌بندی: ادامه
هفدهم	بصری سازی داده‌ها و مدل‌های گرافیکی
روش‌های آموزشی (نحوه ارائه درس): آموزش از طریق سخنرانی و اسلاید، طرح پرسش در هر جلسه	
وسایل آموزشی: ویدئو، پروژکتور، وایت‌برد و مایک	
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 (لطفاً سهم هر روش بر حسب درصد نوشته شود): روش‌های ارزیابی در طول دوره: فعالیت‌های کلاسی (۱۵٪)، آزمون کتبی میان‌ترم (۲۵٪) روش‌های ارزیابی در انتهای دوره: آزمون کتبی پایان‌ترم (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; ۲۰۲۳. ۲. Gareth J, Daniela W, Trevor H, Robert T. An introduction to statistical learning: with applications in R: Springer; ۲۰۱۳. ۳. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer; ۲۰۱۷.	