

بسمه تعالی



دانشگاه علوم پزشکی و خدمات بهداشتی درمانی استان همدان

معاونت آموزشی دانشگاه

مرکز مطالعات و توسعه آموزش علوم پزشکی

فرم طرح درس پایه

مشخصات درس و مدرس

- عنوان درس : اصول و روش‌های داده‌کاوی
- نام و نام خانوادگی مسئول درس: مریم فرهادیان (۲)
- نام و نام خانوادگی مدیر گروه: جواد فردمال
- نوع و میزان واحد به تفکیک: ۲ واحد نظری
- رشته و مقطع تحصیلی دانشجو: کارشناسی ارشد آمارزیستی
- زمان درس: نیمسال اول ۱۴۰۳-۱۴۰۴
- مکان آموزش: دانشکده بهداشت

شماره جلسه	سرفصل (عنوان) جلسه	هدف/اهداف رفتاری	حیطه یادگیری	مدت زمان	وسایل کمک آموزشی	روش ارزشیابی
جلسه ۱	مقدمه‌ای بر داده‌کاوی و یادگیری آماری	<p>- دانشجو بتواند مفهوم داده‌کاوی و یادگیری آماری را تعریف کند.</p> <p>- دانشجو بتواند فرآیند اکتشاف دانش را توضیح دهد.</p> <p>- دانشجو بتواند مفاهیم پایه‌ای نظریه یادگیری آماری را توضیح دهد.</p> <p>- دانشجو بتواند قضیه نبود داده‌کاوی رایگان (No Free Lunch) را تبیین کند.</p> <p>- دانشجو با اصطلاحات و تعاریف اولیه (دیتاست، ویژگی، لیبیل یا برچسب، آموزش مدل، تست مدل و ...) آشنا شود.</p>	دانش (Knowledge)	۱۳۵ دقیقه	اسلایدهای آموزشی، تخته، جزوات	پرسش و پاسخ کلاسی
جلسه ۲	آشنایی با روشهای پیش پردازش داده‌ها	<p>- دانشجو بتواند پیش پردازش و پاکسازی داده‌ها را انجام دهد.</p> <p>- دانشجو بتواند روشهای شناسایی داده‌های پرت را انجام دهد.</p> <p>- دانشجو بتواند روشهای جانمایی مقادیر گم شده را انجام دهد.</p> <p>- دانشجو بتواند روشهای نرمال سازی و استاندارد سازی داده را انجام دهد.</p> <p>- دانشجو بتواند گسسته سازی داده‌ها را انجام دهد.</p>	دانش (Knowledge)	۱۳۵ دقیقه	اسلاید، مثال‌های عددی	تکالیف نوشتاری، مشارکت در بحث
جلسه ۳	آشنایی با تکنیک‌های بصری سازی داده	<p>- دانشجو با انواع روشهای تحلیل اکتشافی و بصری سازی داده‌ها آشنا شود و بتواند آنها را اجرا کند.</p> <p>- دانشجو فرایند شناخت داده‌ها (پراکندگی، توزیع نرمال، تحلیل هیستوگرام، اندازه‌گیری شباهت و عدم شباهت داده‌ها) را درک و اجرا کند.</p> <p>- دانشجو بتواند نقشه گرمایی یا نقشه رنگی را برای یک مجموعه داده ترسیم و اجرا کند.</p>	دانش روان‌حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم‌افزار R یا Python، دیتاست‌های پزشکی	تمرین عملی، آزمون کوتاه
جلسه ۴	آشنایی با روشهای مختلف برش داده	<p>- دانشجو با روشهای مختلف برش داده آشنا و بتواند آنها را اجرا کند.</p> <p>- روش تقسیم داده‌ها به دو مجموعه Train و Test</p>	دانش روان‌حرکتی (Knowledge)	۱۳۵ دقیقه	نرم‌افزار R یا Python، دیتاست‌های پزشکی	تمرین عملی، آزمون کوتاه
جلسه ۵	معرفی الگوریتم‌های مختلف داده‌کاوی	<p>- دانشجو بتواند تفاوت بین یادگیری با نظارت و بدون نظارت را مقایسه کند.</p> <p>- دانشجو بتواند تفاوت الگوریتم‌های مختلف مانند خوشه‌بندی، طبقه‌بندی، استخراج الگوها درک و مقایسه کند</p> <p>- دانشجو مسائل طبقه بندی دودویی binary، چند کلاسه Multi Class، چند برچسب Multi Label درک و تفاوت آنها را بداند.</p> <p>- دانشجو بتواند مفهوم داده‌ی نامتوازن Imbalance در مسایل طبقه بندی را درک و استراتژی‌ها مختلف مواجهه با داده‌های نامتعادل را اجرا کنید.</p>	دانش روان‌حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم‌افزار R یا Python، دیتاست‌های پزشکی	تمرین عملی، آزمون کوتاه

شماره جلسه	سرفصل (عنوان) جلسه	هدف/اهداف رفتاری	حیطه یادگیری	مدت زمان	وسایل کمک آموزشی	روش ارزشیابی
جلسه ۶	طبقه‌بندی و ارزیابی طبقه‌بندی‌ها	<ul style="list-style-type: none"> - دانشجو بتواند روش‌های طبقه‌بندی مختلف را نام ببرد و مقایسه کند. - دانشجو بتواند مدل رگرسیون لجستیک برای طبقه‌بندی را اجرا کند. - دانشجو بتواند تکنیک‌های کلاسیک انتخاب متغیر از جمله روش گام به گام را اجرا کند. - دانشجو بتواند متریک‌های ارزیابی مدل‌های طبقه‌بندی را محاسبه کند. 	دانش روان‌حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم‌افزار R یا Python، دیتاست‌های پزشکی	تمرین عملی، آزمون کوتاه
جلسه ۷	مدل‌سازی با نظارت: روش‌های رگرسیونی	<ul style="list-style-type: none"> - دانشجو بتواند مدل‌های رگرسیونی خطی و غیرخطی را توضیح داده و پیاده‌سازی کند. - دانشجو با مدل‌های رگرسیونی ریج، لاسو و الاستیک نت آشنا و بتواند آنها را اجرا کند. - دانشجو با انواع متریک‌های ارزیابی مدل‌های رگرسیونی آشنا و بتواند آنها را اجرا کند. 	دانش روان‌حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم‌افزار R یا Python، دیتاست‌های پزشکی	تمرین عملی، ارزیابی پروژه کوچک
جلسه ۸	مدل‌سازی بدون نظارت: خوشه‌بندی	<ul style="list-style-type: none"> - دانشجو بتواند مفاهیم خوشه‌بندی را تعریف کند. - دانشجو بتواند الگوریتم‌های K-Means و سلسله‌مراتبی، K-Median، K-medoids را پیاده‌سازی کند. - دانشجو بتواند نتایج خوشه‌بندی را تحلیل کند. - دانشجو بتواند الگوریتم KNN را پیاده‌سازی کند. - دانشجو بتواند معیارهای جینی و انتروپی را محاسبه کند. 	دانش روان‌حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم‌افزارهای تحلیل داده، دیتاست‌های پزشکی	تمرین عملی، گزارش تحلیلی
جلسه ۹	مدل‌سازی بدون نظارت: خوشه‌بندی	<ul style="list-style-type: none"> - دانشجو با معیارهای مختلف ارزیابی روش‌های خوشه‌بندی آشنا شود و آنها را اجرا کند. - آشنایی و اجرای معیارهای درونی و بیرونی ارزیابی - آشنایی و اجرای روش‌های مختلف تعیین تعداد خوشه 				
جلسه ۱۰	تحلیل الگوهای مکرر و قوانین انجمنی	<ul style="list-style-type: none"> - دانشجو بتواند مفاهیم الگوهای مکرر و قوانین انجمنی را توضیح دهد. - دانشجو بتواند الگوریتم Apriori را پیاده‌سازی کند. - دانشجو بتواند نتایج را در زمینه پزشکی تفسیر کند. 	دانش روان‌حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم‌افزارهای داده‌کاوی، دیتاست‌های مرتبط	تمرین عملی، ارزیابی پروژه کوچک

شماره جلسه	سرفصل (عنوان) جلسه	هدف/اهداف رفتاری	حیطه یادگیری	مدت زمان	وسایل کمک آموزشی	روش ارزشیابی
جلسه ۱۱	روش های کاهش بعد و کاهش داده ها	<p>- دانشجو بتواند چالش های داده های با ابعاد بالا را توضیح دهد.</p> <p>- دانشجو بتواند تفاوت بین کاهش بعد و کاهش داده را توضیح دهد.</p> <p>- دانشجو بتواند تفاوت روش های انتخاب ویژگی و استخراج ویژگی را درک کند.</p> <p>- دانشجو بتواند الگوریتم های کاهش داده مانند تحلیل مولفه های اصلی و تحلیل عاملی را اجرا کند.</p> <p>- دانشجو با الگوریتم بروتا آشنا و بتواند آنرا را اجرا کند.</p>	دانش (Knowledge) روان حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم افزارهای تحلیل داده، دیتاست های پزشکی	تمرین عملی، گزارش تحلیلی
جلسه ۱۲	روشهای درخت تصمیم	<p>- دانشجو با الگوریتمهای مختلف درخت تصمیم آشنا شود و آنها را اجرا کند.</p> <p>- الگوریتم ID3</p> <p>- الگوریتم C4.5</p> <p>- الگوریتم CART</p> <p>- الگوریتم جنگل تصادفی</p>	دانش (Knowledge) روان حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم افزار R یا Python، دیتاست های پزشکی	تمرین عملی، ارزیابی پروژه کوچک
جلسه ۱۳	الگوریتم بردار پشتیبان	<p>- دانشجو با الگوریتم بردار پشتیبان خطی و غیر خطی و مفهوم کرنل آشنا شود.</p> <p>- دانشجو بتواند مدل های غیر خطی مانند SVM را پیاده سازی کند.</p> <p>- دانشجو بتواند کاربرد هر مدل را در مسائل پزشکی تحلیل کند.</p>	دانش (Knowledge) روان حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم افزار R یا Python، دیتاست های پزشکی	تمرین عملی، مشارکت در بحث
جلسه ۱۴	ارائه پروژه های عملی	<p>- دانشجو پروژه عملی خود را روی یک مجموعه دیتا واقعی ارائه و اجرا کند</p>	دانش (Knowledge) روان حرکتی (Psychomotor)	۱۳۵ دقیقه	نرم افزار R یا Python، دیتاست های پزشکی	تمرین عملی، مشارکت در بحث

توضیحات تکمیلی:

- روش های تدریس: استفاده از ترکیبی از سخنرانی، بحث گروهی، تمرین های عملی و پروژه های تحقیقاتی.
- وسایل کمک آموزشی: اسلایدهای آموزشی، نرم افزارهای تخصصی (R, Python, WEKA, TensorFlow) و غیره، دیتاست های واقعی در حوزه پزشکی و بهداشت.
- روش های ارزشیابی: شامل تکالیف هفتگی، تمرین های عملی، آزمون های کوتاه، پروژه های کوچک، مشارکت در بحث های کلاسی، و پروژه نهایی.

- تناسب با دوره دکتری آمار زیستی: تمامی سرفصل‌ها و اهداف بر اساس برنامه آموزشی ارائه‌شده و با تمرکز بر کاربردهای عملی در حوزه آمار زیستی و علوم پزشکی طراحی شده‌اند.
- پروژه نهایی: دانشجویان در طول دوره بر روی یک پروژه تحقیقاتی مرتبط با داده‌کاوی در آمار زیستی کار می‌کنند و در جلسه آخر نتایج خود را ارائه می‌دهند.

شیوه نمره دهی:

نوع ارزشیابی	تاریخ	ابزار ارزشیابی ^۱	میزان امتیاز از کل
کوئیز (۲ بار در طی ترم)	جلسه ۹ و جلسه ۱۲	امتحان کتبی	۲۵
ارائه پروژه و تمرین	--	--	۳۵
امتحان پایان ترم	طبق برنامه آموزش	امتحان	۳۵
حضور فعال در کلاس	--	حضور و مشارکت از طریق پاسخگویی و شرکت در بحث	۵
مجموع			۱۰۰

منابع:

1. Hastie T, Tibshirani R, and Friedman J. The Elements of Statistical Learning: data mining, inference and prediction. Springer, Last Edition.
2. Goodfellow I, Bengio Y, Courville A. Deep Learning (Adaptive Computation and Machine Learning series). MIT press, Last Edition.
3. J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, Elsevier Inc., Last Edition.
4. C. C. Aggarwal, Data Mining: The Textbook, Springer, Last Edition.
5. M. J. Zaki, W. Meira. The Handbook of Data Mining, Cambridge University Press, Last Edition.
6. C. C. Aggarwal, Data Streams: Models and Algorithms, Springer, Last Edition.
7. Bifet, Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams, IOS Press, Last Edition
8. Canlas RD. Data Mining in Healthcare: Current Applications and Issues. Last Edition.
9. Llatas, C.Fernández, G.Gómez, J.Miguel. Data Mining in Clinical Medicine, Springer. Last Edition.

^۱ ابزار ارزشیابی می‌تواند مواردی مانند آزمون تشریحی، سوالات کوتاه پاسخ، سوالات کامل کردنی، MCQS، پروژه، آسکی و... باشد.