

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 : روش‌های داده‌کاوی و یادگیری آماری
- نام و نام خانوادگی مسئول درس: مریم فرهادیان (۱/۵)، جواد فردمال (۱)، لیلی تاپاک (۰/۵)
- نام و نام خانوادگی مدیر گروه: جواد فردمال
- نوع و میزان واحد به تفکیک: ۳ واحد نظری
- رشته و مقطع تحصیلی دانشجو: دکتری آمارزیستی
- زمان درس: نیمسال دوم 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	<ul style="list-style-type: none"> - دانشجو بتواند مفهوم داده‌کاوی و یادگیری آماری را تعریف کند. - دانشجو بتواند فرآیند اکتشاف دانش را توضیح دهد. - دانشجو بتواند تفاوت بین یادگیری با نظارت و بدون نظارت را مقایسه کند. - دانشجو بتواند مفاهیم پایه‌ای نظریه یادگیری آماری را توضیح دهد. - دانشجو با اصطلاحات و تعاریف اولیه (دیتاست، ویژگی، لیبیل یا برچسب، آموزش مدل، تست مدل و ...) آشنا شود. 	دانش (Knowledge)	۱۳۵ دقیقه	اسلایدهای آموزشی، تخته، جزوات	پرسش و پاسخ کلاسی
جلسه ۲	نظریه یادگیری آماری	<ul style="list-style-type: none"> - دانشجو بتواند پیش پردازش و پاکسازی داده‌ها را انجام دهد. - دانشجو بتواند روش‌های مختلف اعتبارسنجی متقابل را اجرا کند. - دانشجو بتواند مفهوم بایاس و واریانس را تحلیل کند. - دانشجو بتواند قضیه نبود داده‌کاوی رایگان (No Free Lunch) را تبیین کند. 	دانش (Knowledge)	۱۳۵ دقیقه	اسلاید، مثال‌های عددی	تکالیف نوشتاری، مشارکت در بحث
جلسه ۳	طبقه‌بندی و ارزیابی طبقه‌بندی‌ها	<ul style="list-style-type: none"> - دانشجو بتواند روش‌های طبقه‌بندی مختلف را نام ببرد و مقایسه کند. - دانشجو مسائل طبقه‌بندی دودویی binary، چند کلاسه Multi Class، چند برچسب Multi Label درک و تفاوت آن‌ها را بداند. - دانشجو بتواند متریک‌های ارزیابی مدل‌های طبقه‌بندی را محاسبه کند. - دانشجو بتواند مفهوم داده‌ی نامتوازن Imbalance در مسایل طبقه‌بندی را درک و استراتژی‌ها مختلف مواجهه با داده‌های نامتعادل را اجرا کند. 	روان‌حرکتی (Knowledge) ، دانش (Psychomotor)	۱۳۵ دقیقه	نرم‌افزار R یا Python، دیتاست‌های پزشکی	تمرین عملی، آزمون کوتاه
جلسه ۴	مدل‌سازی با نظارت: روش‌های پیش‌بینی	<ul style="list-style-type: none"> - دانشجو بتواند مدل‌های رگرسیونی خطی و غیرخطی را توضیح داده و پیاده‌سازی کند. - دانشجو بتواند الگوریتم KNN را اجرا کند. - دانشجو بتواند دقت مدل‌های پیش‌بینی را ارزیابی کند. 	دانش (Knowledge) ، روان‌حرکتی	۱۳۵ دقیقه	نرم‌افزار R یا Python، دیتاست‌های پزشکی	تمرین عملی، ارزیابی پروژه کوچک

شماره جلسه	سرفصل (عنوان) جلسه	هدف/اهداف رفتاری	حیطه یادگیری	مدت زمان	وسایل کمک آموزشی	روش ارزشیابی
جلسه ۵	مدل سازی بدون نظارت: خوشه بندی	<ul style="list-style-type: none"> - دانشجو بتواند مفاهیم خوشه بندی را تعریف کند. - دانشجو بتواند الگوریتم های K-Means و سلسله مراتبی، K-Median، K-medoids را پیاده سازی کند. - دانشجو با معیارهای مختلف ارزیابی روشهای خوشه بندی آشنا شود و آنها را اجرا کند. - دانشجو بتواند نتایج خوشه بندی را تحلیل کند. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	نرم افزارهای تحلیل داده، دیتاست های پزشکی	تمرین عملی، گزارش تحلیلی
جلسه ۶	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 پیاده سازی کند. - دانشجو بتواند نتایج را در زمینه پزشکی تفسیر کند. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	نرم افزارهای داده کاوی، دیتاست های مرتبط	تمرین عملی، ارزیابی پروژه کوچک
جلسه ۷	الگوهای تصمیم گیری و مدل های خطی و غیر خطی	<ul style="list-style-type: none"> - دانشجو بتواند مدل های خطی مانند Logistic Regression را توضیح دهد. - دانشجو بتواند مدل های غیرخطی مانند SVM را پیاده سازی کند. - دانشجو بتواند کاربرد هر مدل را در مسائل پزشکی تحلیل کند. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	اسلاید، نرم افزارهای داده کاوی	تمرین عملی، مشارکت در بحث
جلسه ۸	شبکه های عصبی و یادگیری عمیق	<ul style="list-style-type: none"> - دانشجو با الگوریتم گرادیان کاهشی آشنا شود. - دانشجو بتواند ساختار شبکه های عصبی را توضیح دهد. - دانشجو بتواند یک شبکه عصبی ساده را پیاده سازی کند. - دانشجو بتواند کاربردهای یادگیری عمیق در پزشکی را نام ببرد. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	نرم افزارهای یادگیری عمیق (مانند TensorFlow یا Keras)	تمرین عملی، ارائه کلاسی
جلسه ۹	روشهای درخت تصمیم	<ul style="list-style-type: none"> - دانشجو با الگوریتمهای مختلف درخت تصمیم آشنا شود و آنها را اجرا کند. - الگوریتم ID3 - الگوریتم C4.5 - الگوریتم CART - الگوریتم جنگل تصادفی 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	نرم افزار R یا Python، دیتاست های پزشکی	تمرین عملی، ارزیابی پروژه کوچک
جلسه ۱۰	مفهوم یادگیری در داده های با بعد بالا	<ul style="list-style-type: none"> - دانشجو بتواند چالش های داده های با ابعاد بالا را توضیح دهد. - دانشجو بتواند تفاوت روش های انتخاب ویژگی و استخراج ویژگی را درک کند - دانشجو بتواند روش های انتخاب ویژگی را مقایسه کند. - دانشجو بتواند تکنیک های کاهش بعد مانند PCA را اجرا کند. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	اسلاید، نرم افزارهای داده کاوی	تکالیف عملی، مشارکت در بحث

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	<ul style="list-style-type: none"> - دانشجو بتواند تفاوت بین کاهش بعد و کاهش داده را توضیح دهد. - دانشجو بتواند الگوریتم های کاهش داده مانند Sampling را اجرا کند. - دانشجو بتواند کاربردهای عملی کاهش بعد در آمار زیستی را تحلیل کند. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	نرم افزارهای تحلیل داده، دیتاست های پزشکی	تمرین عملی، گزارش تحلیلی
جلسه ۱۲	خوشه بندی پیشرفته و تحلیل داده های پرت	<ul style="list-style-type: none"> - دانشجو بتواند الگوریتم های خوشه بندی مبتنی بر تراکم و الگوریتم EM را اجرا کند. - دانشجو بتواند داده های پرت را شناسایی و تحلیل کند. - دانشجو بتواند روش های خوشه بندی داده های با ابعاد بالا را توضیح دهد. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	نرم افزارهای داده کاوی، دیتاست های پزشکی	تمرین عملی، گزارش تحلیلی
جلسه ۱۳	قواعد تصمیم گیری: بیزی، انترویی، نزدیک ترین همسایه و غیره	<ul style="list-style-type: none"> - دانشجو بتواند اصول تصمیم گیری بیزی را توضیح دهد. - دانشجو بتواند الگوریتم KNN را پیاده سازی کند. - دانشجو بتواند معیارهای جینی و انترویی را محاسبه کند. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	اسلاید، نرم افزارهای داده کاوی	تمرین عملی، آزمون کوتاه
جلسه ۱۴	انبار داده و استخر داده	<ul style="list-style-type: none"> - دانشجو بتواند مفاهیم انبار داده (Data Warehouse) و استخر داده (Data Lake) را توضیح دهد. - دانشجو بتواند کاربردهای آن ها در سازمان های بهداشتی را نام برد. - دانشجو بتواند مراحل طراحی یک انبار داده را شرح دهد. 	دانش (Knowledge)	۱۳۵ دقیقه	اسلاید، موارد مطالعاتی	پرسش و پاسخ، ارائه کلاسی
جلسه ۱۵	استخراج مجموعه های زمانی و داده های جریانی	<ul style="list-style-type: none"> - دانشجو بتواند مفاهیم سری های زمانی را توضیح دهد. - دانشجو بتواند الگوریتم های داده کاوی در داده های جریانی را نام برد. - دانشجو بتواند الگوهای پرتکرار در داده های جریانی را شناسایی کند. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	اسلاید، نرم افزارهای تخصصی	تمرین عملی، ارائه کلاسی
جلسه ۱۶	داده کاوی در فضایی و گراف کاوی	<ul style="list-style-type: none"> - دانشجو بتواند مفاهیم داده های فضایی و فضایی-زمانی را توضیح دهد. - دانشجو بتواند روش های گراف کاوی را شرح دهد. - دانشجو بتواند زیرگراف های پرتکرار را در داده های گرافی شناسایی کند. 	دانش روان حرکتی (Knowledge) (Psychomotor)	۱۳۵ دقیقه	اسلاید، نرم افزارهای گراف کاوی	تمرین عملی، پروژه نهایی

توضیحات تکمیلی:

- روش‌های تدریس: استفاده از ترکیبی از سخنرانی، بحث گروهی، تمرین‌های عملی و پروژه‌های تحقیقاتی.
- وسایل کمک آموزشی: اسلایدهای آموزشی، نرم‌افزارهای تخصصی (TensorFlow, WEKA, Python, R) و غیره، دیتاست‌های واقعی در حوزه پزشکی و بهداشت.
- روش‌های ارزشیابی: شامل تکالیف هفتگی، تمرین‌های عملی، آزمون‌های کوتاه، پروژه‌های کوچک، مشارکت در بحث‌های کلاسی، و پروژه نهایی.
- تناسب با دوره دکتری آمار زیستی: تمامی سرفصل‌ها و اهداف بر اساس برنامه آموزشی ارائه شده 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.