



فرم معرفی درس روش های داده کاوی و یادگیری آماری – نیمسال دوم ۱۴۰۴-۱۴۰۵
دانشکده بهداشت، گروه آموزشی اپیدمیولوژی و آمار زیستی
رشته آمار زیستی، مقطع PhD

شماره درس: ۴۱۳۶۱۰۱۰
روز و ساعت برگزاری: سهشنبه ۱۱-۰۸
درس های پیش نیاز: ندارد
نام مسوول درس: دکتر کامران مهربانی
آدرس دفتر: دانشکده بهداشت، گروه آمارزیستی و اپیدمیولوژی
تعداد و نوع واحد (نظری/عملی): ۳ واحد نظری
محل برگزاری: دانشکده بهداشت – سایت نور
تلفن و روزهای تماس: ۳۷۹۲۳۱۷۴ _ سهشنبه ۱۴-۱۶
آدرس Email: k.mehrabani@hlth.mui.ac.ir

هدف کلی درس:

توانمندسازی دانشجویان در درک نظری و به کارگیری عملی روش های داده کاوی و الگوریتم های یادگیری ماشین به منظور استخراج الگوهای اساسی، تحلیل مجموعه داده های بزرگ، و تفسیر نتایج در پژوهش های علوم پزشکی.

اهداف اختصاصی درس:

۱. مدیریت مجموعه داده های بزرگ، پالایش و آماده سازی داده ها.
۲. آشنایی و به کارگیری شیوه های مدیریت مجموعه داده های با پاسخ نامتوازن در مدل های پیش بینی.
۳. آشنایی و استفاده از شاخص های ارزیابی عملکرد الگوریتم های پیش بینی.
۴. شیوه های انتخاب متغیرهای اثرگذار، بررسی تفاوت دیدگاه روش های Embedded و Wrapper.
۵. اجرا و تفسیر روش های کاهش بعد مانند مولفه های اصلی /عامل و LDA.
۶. به کارگیری الگوریتم های رایج خوشه بندی و شناخت معیارهای فاصله.
۷. تفسیر معیارهای درونی و بیرونی ارزیابی کیفیت خوشه بندی / تعیین تعداد خوشه ها.
۸. یادگیری گام های تشکیل درخت تصمیم و تفسیر قاعده های استخراج شده.
۹. تحلیل مفهوم های آنروپی و شاخص جینی در ساخت درخت تصمیم.
۱۰. شناخت روش های متداول مبتنی بر درخت، مانند جنگل تصادفی.
۱۱. آشنایی با شاخص های اهمیت متغیرها، نحوه محاسبه و تفسیر آن ها.
۱۲. آشنایی با شیوه های ارتقای روش های یادگیری ماشین با بهره گیری از رویکردهای آماری.
۱۳. اجرا و تحلیل روش های تنبل، مانند k-نزدیک ترین همسایگی.
۱۴. تشریح مبانی ماشین بردار پشتیبان از دیدگاه هندسی و آماری.
۱۵. مدل سازی و تفسیر شبکه های عصبی پایه.
۱۶. به کارگیری نرم افزار R در تحلیل داده های واقعی.
۱۷. انجام حداقل یک پروژه کامل داده کاوی از پالایش و پیش پردازش مجموعه داده تا تحلیل و گزارش نهایی.
۱۸. تفسیر کاربردی نتیجه های مدل ها در چارچوب پژوهش های علوم پزشکی.
۱۹. معرفی تکنیک های انجام گام ها به صورت پویا.
۲۰. بصری سازی نتایج برای شناخت و مقایسه حجم زیادی از خروجی ها.

منبع‌های اصلی درس:

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
2. Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Elsevier.
3. Zaki, M. J., & Meira, W. (2014). Data mining and analysis: Fundamental concepts and algorithms. Cambridge University Press.

منبع‌های فرعی درس: ویدیوها، اسلایدها و موردهای درس داده شده در کلاس.
روش تدریس: سخنرانی، پرسش و پاسخ، بحث گروهی، آزمون و مرور آموزه‌های جلسه‌های پیشین.
مسئولیت‌های فراگیرها: حضور فعال در بحث‌های کلاسی، پاسخ به تمرین‌های کلاس، انجام تکلیف‌ها، مرور آموزه‌ها پس از هر جلسه و بحث و بررسی آن‌ها در جلسه بعد.

نحوه ارزشیابی دانشجو و بارم مربوط به هر ارزشیابی:

الف) در طول دوره:

- ارزشیابی تکوینی شامل فعالیت کلاسی | ۳ نمره.
- ارایه مبحث کلاسی | ۵ نمره.

ب) پایان دوره:

- پروژه نهایی | ۶ نمره
- امتحان کتبی | ۶ نمره

سیاست مسئول دوره در مورد برخورد با غیبت و تاخیر دانشجو در کلاس درس:
حضور دانشجویان طبق قانون آموزش سنجیده خواهد شد.

جدول زمان بندی ارائه برنامه درس

ردیف	تاریخ	ساعت	عنوان	مدرس	آمادگی لازم دانشجویان قبل از شروع کلاس
۱	۱۴۰۴/۱۲/۰۵		ارایه نمای کلی درس و مشکل‌های ناشی از داده نامتوازن (به صورت مجازی)		-
۲	۱۴۰۴/۱۲/۱۲		شیوه‌های مدیریت مجموعه‌داده‌های با پاسخ نامتوازن (به صورت مجازی)		
۳	۱۴۰۴/۱۲/۱۹		انتخاب متغیرهای اثرگذار، دیدگاه‌های Embedded و Wrapper (به صورت مجازی)		
۴	۱۴۰۴/۱۲/۲۶	۱۱-۰۸	مجموعه‌داده‌های بزرگ، پالایش و آماده سازی آن‌ها (به صورت مجازی)	دکتر مهربانی	مرور آموزه‌های جلسه قبل و منبع درس
۵	۱۴۰۵/۰۱/۱۸		تشکیل درخت تصمیم و تفسیر قاعده‌های استخراج شده		
۶	۱۴۰۵/۰۱/۲۵		روش‌های متداول مبتنی بر درخت، مانند جنگل تصادفی		
۷	۱۴۰۵/۰۲/۰۱		شاخص‌های اهمیت متغیرها، نحوه محاسبه و تفسیر آن‌ها		
۸	۱۴۰۵/۰۲/۰۸		روش‌های کاهش بعد مانند مولفه‌های اصلی/عامل و LDA		
۹	۱۴۰۵/۰۲/۱۵		مفهوم و تکنیک‌های خوشه‌بندی، معیارهای فاصله		
۱۰	۱۴۰۵/۰۲/۲۲		ارزیابی خوشه‌بندی و کاربرد در علوم پزشکی		
۱۱	۱۴۰۵/۰۲/۲۹		شیوه‌های ارتقای روش‌های یادگیری ماشین با رویکردهای آماری		

		معرفی تکنیک‌های انجام گام‌ها به صورت پویا	۱۴۰۵/۰۳/۰۵	۱۲
		بصری‌سازی نتایج برای شناخت و مقایسه حجم زیادی از خروجی‌ها	۱۴۰۵/۰۳/۱۲	۱۳
		KNN و روش‌های LAZY	۱۴۰۵/۰۳/۱۹	۱۴
		ماشین بردار پشتیبان	۱۴۰۵/۰۳/۲۶	۱۵
		شبکه‌های عصبی پایه (Perceptron, MLP)	۱۴۰۵/۰۴/۰۲	۱۶
		مرور مبحث‌ها و رفع اشکال پیش از آزمون	۱۴۰۵/۰۴/۰۹	۱۷

تاریخ امتحان پایان ترم: طبق تقویم آموزش

سایر تذکرات مهم برای دانشجویان:

- در صورت غیبت در آزمون، آزمون دوباره گرفته نخواهد شد.
- برای دریافت نمره فعالیت کلاسی باید در مباحث کلاسی مشارکت داشته باشید.

مهربانی