

کد درس: ۱۴

نام درس: تحلیل داده‌های سلامت

پیش‌نیاز یا همزمان: ندارد

تعداد واحد: ۲ واحد (۱ واحد نظری - ۱ واحد عملی)

نوع واحد: نظری - عملی

هدف کلی درس: آشنایی دانشجویان با انواع داده‌ها، مفاهیم و تکنیک‌های داده کاوی، اصول کاربردی در پروژه‌های داده محور، پیاده‌سازی روش‌های داده کاوی و ارزشیابی آنها و مقدمه‌ای بر چگونگی مواجهه با داده‌های حجیم و یادگیری عمیق. شرح درس: دانشجو در این درس ضمن آشنایی با مفاهیم داده کاوی و تکنیک‌های آن، مراحل انجام پروژه‌های داده کاوی، چگونگی آماده‌سازی داده‌ها برای تحلیل، روش‌های با ناظر و بدون ناظر، توانایی تحلیل داده‌ها و استخراج دانش از میان گونه‌های متفاوت داده‌ها در حوزه فناوری اطلاعات سلامت را کسب می‌کند.

رئوس مطالب نظری (۱۷ ساعت):

- تعریف داده و داده کاوی و بررسی انواع داده‌های حوزه سلامت
- تفاوت داده کاوی و آمار و تفاوت‌های داده، اطلاعات و دانش در حوزه سلامت
- مراحل انجام پروژه‌های داده کاوی
- پیش پردازش داده‌ها (چگونگی آماده‌سازی داده‌های حوزه سلامت برای تحلیل، مصورسازی داده‌ها، چگونگی مواجهه با داده‌ها و مقدارهای ناموجود)
- تجمیع داده‌ها به همراه تفاوت‌ها و شباهت‌های پایگاه، مخزن و انبار داده
- کلاسه‌بندی و خوشه‌بندی داده‌ها به همراه استخراج الگوهای موجود در داده‌ها، استفاده از تکنیک‌های کاهش حجم داده‌ها مانند نمونه‌گیری، روش‌های کاهش ابعاد مانند تحلیل مولفه اصلی (PCA)
- روش‌های با ناظر
 - o بررسی روش‌های کلاسه‌بندی مانند: درخت تصمیم، ماشین بردار پشتیبان، رده‌بندی مبتنی بر قواعد، نزدیکترین نمونه، رگرسیون و انواع آن
 - o ماتریس درهم‌ریختگی، چگونگی ارزیابی روش‌های رده‌بندی و بررسی متریک‌هایی چون صحت، دقت و بازخوانی (حساسیت) روش‌های بدون ناظر
 - o الگوهای مکرر، قواعد انجمنی، بررسی متریک‌هایی چون پشتیبان و اطمینان
 - o خوشه‌بندی و انواع آن، الگوریتم k-means، و نسخه‌های مختلف آن، روش‌های خوشه‌بندی سلسله مراتبی
- کار با نرم افزارهای داده کاوی مانند رپیدماینر روی داده‌های عددی و غیرعددی حوزه سلامت (در صورت وجود پیش زمینه در دانشجویان می‌توان از نرم افزارهای برنامه‌نویسی مانند متلب به جای رپیدماینر استفاده کرد).
- مقدمه‌ای بر کار با داده‌های حجیم:
 - o خصوصیات داده‌های بزرگ (حجم، تنوع، سرعت تولید) و چالش‌های کار با آن به همراه دلیل عدم کارایی روش‌های متداول در حوزه سلامت
 - o آشنایی با نمونه‌هایی از داده‌های بزرگ
 - o مقدمه‌ای بر تکنیک‌های پیشرفته مصورسازی و تحلیل داده‌های بزرگ
- مقدمه‌ای بر یادگیری عمیق
 - o شبکه‌ها و روش‌های عصبی متداول مانند پرسپترون و روش بازگشت به عقب
 - o دلیل ناکارآمدی روش‌های متداول در شبکه‌های عصبی
 - o خصوصیات روش‌های عمیق و چالش‌های کار با این روش‌های یادگیری به همراه مقدمه‌ای بر چند روش پایه



رئوس مطالب عملی (۳۴ ساعت):

- انجام پروژه‌های داده کاوی
- بکارگیری اصول پیش پردازش داده‌ها (چگونگی آماده‌سازی داده‌های حوزه سلامت برای تحلیل، مصورسازی داده‌ها، چگونگی مواجهه با داده‌ها و مقدرهای ناموجود)
- تجمیع داده‌ها با توجه به تفاوت‌ها و شباهت‌های پایگاه، مخزن و انبار داده، کلاسه‌بندی و خوشه‌بندی داده‌ها و استخراج الگوهای موجود در داده‌ها
- استفاده از تکنیک‌های کاهش حجم داده‌ها مانند نمونه‌گیری، روش‌های کاهش ابعاد مانند تحلیل مولفه اصلی (PCA)

منابع اصلی درس:

۱. شبکه‌های عصبی و یادگیری ژرف، مهدی اسماعیلی و فائزه اکبری، تهران: انتشارات نیاز دانش، آخرین ویرایش
2. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, Last Edition.
3. Masmoudi M, Jarboui B, Siarry P, Artificial Intelligence and Data Mining in Healthcare, Springer International Publishing, Last Edition.
4. Bishop C M, Pattern Recognition and Machine Learning, Springer, Last Edition.
5. Zhang A, Lipton Z C, Li M, Smola A J, Dive into Deep Learning, Last Edition.

شیوه ارزشیابی دانشجو:

- حیطه شناختی: ارزشیابی دانشجو در اواسط و پایان دوره به صورت کتبی انجام می‌شود.
- حیطه روانی- حرکتی: ارزشیابی دانشجو به صورت پروژه محور است. بدین ترتیب که دانشجو پس از انتخاب و جمع‌آوری مجموعه داده‌های خاص، با استفاده از ابزارهای داده کاوی مناسب، چند روش داده کاوی را بر روی داده‌ها اجرا می‌کند.

